DOCUMENT RESUME

ED 149 493 EC 103 453

AUTHOR	Wessitsh, Alexander
TITLE	Optimizing Habilitation Services in the Public Schools: Summative Evaluation Report.
INSTITUTION	Kansas Univ., Lawrence. School of Education.
SPONS AGENCY	Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C. Div. of Training Programs.
BUREAU NO	451AH50059
PUB DATE	Nov 76
GRANT	G007501808
NOTE	200p.
EDRS PRICE	MF-$0.83 HC-$10.03 Plus Postage.
DESCRIPTORS	Educational Objectives; Exceptional Child Research; Federal Government; Grants; *Learning Disabilities; *Material Development; Performance Based Teacher Education; Pilot Projects; Post Secondary Education; *Research Projects; *Special Education Teachers; *Summative Evaluation; *Teacher Education; Teacher Programs
IDENTIFIERS	Research Reports

ABSTRACT

Described is an evaluation report, the Habilitation Personnel Training Project, designed to develop a model and materials for training secondary special education personnel such as classroom teachers and work study coordinators. The objectives of the project, which were carried out by the University of Kansas, are given in the introductory section as well as description of the term "evaluation" as used in the title of the report. It is noted that the project was funded by the Bureau of Education for the Handicapped under Special Programs for Children with Specific Learning Disabilities, Parts D and G of Title VI, the Education for the Handicapped Act, Public Law 91-230. The history and origin of the project are described as well as problems incurred (such as uncontrolled evolution of concepts and failure to control communication and decision making), and solutions that worked or failed to solve problems. Summarized are four models used to carry out the project plans--the content, training, evaluation, and dissemination models. Appendixes provide information on adjunct research, instructor and module evaluation, factor cluster of items, and a listing of factors and "Skree" analysis. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED149493

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SPECIAL PROJECT: OPTIMIZING HABILITATION SERVICES IN THE PUBLIC SCHOOLS

SUMMATIVE EVALUATION REPORT

prepared by:

ALEXANDER WESSITSH, coordinator evaluator

JAMES P. GIRARD, project editor

SHERMAN YACHER, graduate research assistant

edited by:

ALEXANDER WESSITSH

artwork by:

MARIANNE NOTLEY

November, 1976
University of Kansas
Habilitation Personnel Training Project
Dr. Gary M. Clark, Project Director

Department of Special Education
School of Education
Haworth Hall
Lawrence, Kansas 66045

EC103453

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gary M. Clark

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Project No. 451AH50059
Grant No. GOO 7501808

The research reported herein
was performed pursuant to a grant
with the Bureau of Education for the Handicapped,
U.S. Office of Education,
Department of Health, Education and Welfare.
Contractors undertaking such projects
under government sponsorship
are encouraged to express freely their professional judgment
in the conduct of the project.
Points of view or opinions stated
do not, therefore, necessarily represent official positions
of the Bureau of Education for the Handicapped.

Department of Health, Education and Welfare
U.S. Office of Education
Bureau of Education for the Handicapped
Division of Training

Preface

After completing four years of concentrated effort on the Habilitation Personnel Training Project, one would think that something more profound than the Alka Seltzer commercial, "Plop-plop, fizz-fizz, oh, what a relief it is!" would come to mind. However, as an index of the candor which the reader will encounter in this report, perhaps such honest reactions can be reported and be considered appropriate, and that is the way we feel.

One of the reasons we feel relief to have reached the terminal stage of the project is that we recognize that we have accomplished a great deal in an area that has been a pioneer territory. Competency-based teacher education was reaching its zenith during the first two years of this project, and most of what was available in the literature and from resource persons was logic and enthusiasm for the concept. Very little empirical research existed then to shed light on the uncharted course we had committed ourselves to follow. It is good to be through any period which constantly reminds one that there is still so much to learn.

Completion of the day-to-day operations of a project, however, does not provide a complete sense of relief. There is always the knowledge that so much still needs to be done. There is the sense of responsibility for having produced something that was supported by public funds and a feeling of obligation to stand accountable. As director of the project from its inception, I assume full responsibility for the weaknesses and gaps which will be reported in these pages.

For those aspects of our efforts for which we take most pride, I want to credit those who made it possible. First, to those who served in administrative capacities, I want to acknowledge Warren D. Schoonmaker, Stuart E. Schwartz, Jim Girard, Keith Edmonds, Greg Weisenstein, and Alexander Wessitsh. Secondly, the writing and media staff, Norma Blankinship, Keith Edmonds, Karen DeVictor, Marianne Notley, Rod Oelschlager, Lynne Prouty, Bonnie Rosefelt, Linda Thompson, and Sherman Yacher, all contributed immeasurably to the development of the field test modules and deserve considerable credit for the good things to emerge as project products. Third, I wish to express deep appreciation to the graduate assistants who responded so well under intense pressure from their individual degree programs and the demands of the project: Mary Eastwold, Sandra Petak Carleton, Mike Fitzgerald, Marjorie Getz Sheftel, John Kregel, Marcy Leeds, Sabra Ohlsen, Bonnie Rosefelt, Warren D. Schoonmaker, Kathaleen Shaffer, Patricia L. Sitlington, Stuart E. Schwartz, Greg Weisenstein, Diane Wimmer, and Sherman Yacher. Finally, and in the practical sense, most importantly, I want to acknowledge Cris Darmon, Sharon Herrick, Susie Mathewson, and Gayle Yelenik for keeping the process and product going and who, of all the staff, were the only ones who knew what they were doing at all times.

This project was designed to develop a model for training secondary special education personnel, i.e., classroom teachers and work-study coordinators, and some materials to facilitate that model. This was done, but it should be considered only a beginning. Competency-based teacher education, as presented in this report, still retains its logic, but enthusiasm is tempered. There is still no substitute for positive human interaction in the instructional process. Competency-based education which attempts to trade off that positive interpersonal communication for an efficient, but mechanistic, set of objectives and canned activities is doomed to eventual failure in spite of satisfactory post-test scores.

The dissemination of the products of this project await commercial publisher response. Interested users should contact me until distribution procedures are resolved.

Gary M. Clark, Project Director
Habilitation Personnel Training Project
University of Kansas
Lawrence, Kansas 66045

TABLE OF CONTENTS

Introduction 1
  Sources of goals and objectives for HPTP 5
    Governmental 5
    The University of Kansas 6
    HPTP (The Habilitation Personnel Training Project) 7
Actions Taken and Outcomes Resulting 15
  An evaluative history of the Habilitation Personnel Training Project. 18
    The origin of the Project and the birth of the Lattice 19
    Problem number one: The uncontrolled evolution of concepts. 23
    Problem number two: The failure to control communication and decision making. 27
    A solution that failed: The Implementation Model 35
    The solution(s) that worked. 39
    The last two years: A few remaining problems 45
    Conclusions. 49
  The Content Model 52
  The Training Model. 56
    Training Model outcomes. 60
      Instructor Module evaluations 60
        Summary statements 60
        Overview 63
        Converted ratings. 65
        Summary comments 69
        Conclusions. 70
      Module evaluation and achievement data. 71
        Attribution theory 72
        Procedures 75
        Factor analysis. 76
        Module Evaluation reliability. 81
        Combined point biserial correlations 86
        Achievement. 92
        Conclusions. 98

The Evaluation Model 99
Project goals 99
Product goals 110
Summative evaluation. 112
The Dissemination Model. 114
Discussion and recommendations. 125
Appendix A: Adjunct Research 135
Cox-Vargas PPDI analysis of some items from *The World of Work* 136
Methodology 145
Results and conclusions 148
References. 154
Subappendix A 157
1. Demographic data form. 158
2. Letter to experienced subjects 159
Subappendix B 161
Item analysis sample 162
Appendix B. 165
1. Instructor Evaluation. 166
2. Module Evaluation. 168
Appendix C. 173
17 Factor clusters of items. 174
Appendix D. 179
Listing of 54 factors and "Skree" analysis 180
References. 181

LIST OF FIGURES AND TABLES

Figure 1: The Project's Planning Model 17
Figure 2: The Content Model. 53
Figure 3: The Training Model 57
Figure 4: The Formative Evaluation Model 103
Figure 5: The Module Production Process including Formative Evaluation Step 5.0 104
Figure 6: The Evaluation Revision Cycles in Step 5.0 105
Figure 7: Summative Evaluation Model 113
Figure 8: The Dissemination Model. 115

* * *

Table 1: Instructor Module Evaluations Converted to Ratings 64
Table 2: ModuleEx - Combined Point Biserial Example 87
Table 3: Two Pairs of Indices for Preliminary Assessment of Achievement. 95
Table 4: Special Education Student Teachers Who Received Funds from HPTP 101
Table 5: Professions of Registrants at the 1976 CEC Convention. 120
Table 6: Chronological Listing of Project Documents Disseminated 123

INTRODUCTION

This report is subtitled a "summative evaluation report." It is summative in the spirit of the discussion begun by Scriven (1967). It is an "evaluation" report because it will attempt to assign estimates of the worth of HPTP work.

It must be said at this point that there is often confusion between the terms "evaluation" and "research." Confusion over the activities referred to by these terms leads to confusion in expectations. "Evaluation," as noted, refers to the assignment of value or worth to something. Values lie within people and people tend to use them relative to their understanding of events. Consequently, judgements tend to be made relatively (Kohlberg, 1968). For example, to give an ice cream cone to an infant may be evaluated as good, bad or somewhere in between, depending on whether the baby is hungry, fat, or is getting it all over his bib. In addition, evaluation also depends on estimates of the truth of things. Did the man knock the woman down to steal her purse, or was it to save her from a nearby speeding car? The truth about possession of the purse and path of the car must be ascertained before a full evaluation of the man and his actions can be made. Evaluation is based on people's values and on what they believe to be true.

Research is concerned with truth or reality. Research, in its most productive form, is concerned with the empirical truth of theories about the world. However, research can also be used to determine the facts about the path of a car. Typically, there is inadequate evidence to determine, without question, what is or was the case. Not enough witnesses were in a good position to tell the direction and speed of the car. Thus, a detective-researcher cannot tell if the man was justified in knocking the woman down. However, with a number of other kinds of evidence accumulated, a theory or model of what happened might be developed. In short, some overview and integration of information about this incident would be developed and then the detective-researcher could infer what motivations the man had. It is necessary for one to estimate the truth of the situation to be able to evaluate the man.

Similarly with a project, the truth about the activities and products to be evaluated must be arrived at by using a variety of information and integrating it. Then, to complete evaluation, some set of values must be applied to the true picture developed from the integrated information. In practice, this should mean that research activities should precede evaluation activities. As a matter of fact, however, research about the true state of things nevers seems to match the need for evaluation. Either a topic is under theoretical debate, like the nature of intelligence, or the tools to carry out the research are not available, such as the reliability of criterion referenced tests, or perhaps no research has been done yet and little can be done on the budgeted evaluation time, money and talent. For these reasons, when the truth about things to be evaluated is not already available and self evident, research evidence often cannot be mounted to support an evaluation. In these cases the evaluation must take place based on scant facts but with abundant estimates of value available.

Related to these differences between evaluation and research, is a difference between "intellectual" and "social" evaluation. The distinction being drawn here is between technical excellence and praiseworthiness. The label "intellectual" is used because most of the criteria have to do with logically thought out rules of excellence. "Social" is used to refer to criteria expressed mostly in terms of how excellent the solutions are for social problems. A significant difference in mean reading scores of three tenths grade point would be excellent. It could be "intellectually" excellent because it was a statistically significant difference. It could be "socially" excellent because it would indicate an increase in the ability to use the tool of reading for personal and social reasons. Part of the issue focused on in this distincition between "intellectual" and "social" excellence is the sources of values applied. If research science and "intellectual" sources are appealed to, sensitivity to needs may lapse and if "social" sources, well thought out consideration of alternatives may give way to bias. This means that there is, between intellectual and social values, an evaluation to be made. The outcome of this evaluation would not not necessarily be a choice between the two; in fact, the preferable outcome should be a weighing of the two.

By making this "social-intellectual" division in evaluation, evaluation is described as composed of activities related to intellectual and social criteria and as related to research by the necessity for having facts. Evaluation not only depends on the kinds of criteria used and the sources of information, but also on specification of those criteria. Some criteria are written down as goals and objectives. Others are implied and not specific. Some are self-serving and others are for service. Further, in a project where a university teacher education curriculum is being developed, there are several levels of goals and objectives. There are goals and objectives for (1) the project regarding research and curriculum, (2) for the professors who will use the curriculum, (3) for the teacher trainees who are to be educated with the curriculum, and (4) for the pupils who will be served by the teachers. Since there are so many levels and so many possible goals for each level, the specification of goals and objectives can become very complex and confusing. In the other extreme, the goals may be specified for only one level or sprinkled throughout the levels. Such disparate specification would mean inadequate coverage at most levels. Evaluation based on disparate and changing goals is equivocal and perhaps nonsensical.

The essence of these introductory paragraphs has been to develop some ideas necessary to keep in mind while evaluation is done and while one writes or reads an evaluation report. In summary, these ideas are the following. Evaluation is not research, although it depends on having facts that can be developed from research. Evaluation can be intellectual and social. It can be based on technical criteria and on practical criteria. Evaluation is keyed to some form of goals and objectives. Sometimes the goals and objectives are explicit and stated, other times vague or unstated. Goals and objectives for curriculum development span several levels of users of the curriculum. These are key ideas which should clarify ambiguities causing confusion in evaluation.

Sources of Goals and Objectives for HPTP

Governmental

HPTP has been funded through the Bureau of Education for the Handicapped under legislation titled, "Training of Personnel for the Education of the Handicapped" and "Special Programs for Children with Specific Learning Disabilities." These are Parts D and G under Title VI, the Education for the Handicapped Act, of Public Law 91-230. This legislation was designed to provide grants to institutions of higher education to assist them in providing training of professional personnel to conduct training of teachers, in providing training for personnel engaged or preparing to engage in employment as teachers of handicapped children, and in establishing and maintaining scholarships for training personnel engaged in or preparing to engage in employment as teachers of the handicapped. It also provides for grants to carry out a program of research and related purposes relating to the education of children with specific learning disabilities and to carry out a program of professional or advanced training for educational personnel who are teaching, or are preparing to be teachers of, children with specific learning disabilities.

The University of Kansas

Being situated in the Department of Special Education, School of Education at the University of Kansas, also affects HPTP. Section H.1 of the Handbook for Faculty and the Unclassified Staff states, "The University of Kansas encourages research by its faculty." "Research" in this context is taken to mean more than scientific manipulation of data because it covers literacy and other artistic endeavor as well as science. Section H.2 goes on to state, "University policy concerning research is derived from the University's general goals. Among the University's most basic goals are the development, use, and dissemination of knowledge." Included with these basic goals is the goal of ". . . constant submission of all knowledge, new and old, for reappraisal by anyone concerned."

A further part of the process of disseminating knowledge is to disseminate the techniques of reappraisal. The spirit of these goals is not unique to the University of Kansas. In fact, it is so pervasive that statements like those presented above serve primarily to codify what in reality is already practiced.

The practices of reappraisal encourage argumentativeness. Overuse of the principle can impede agreement on practical means of accomplishing goals. The result would be that the one set of principles gets in the way of the other set when neither should be superordinate. Thus, the content of HPTP modules was intentionally subjected to review and revision, while at the same time efforts were carried on to make the content fixed and settled for module production. Occasionally an argumentative character to reappraisal would develop. Such an outgrowth delayed achievement of production goals. In short, the goals of the University, in fact probably of most U.S. universities, mediate the goals of the projects funded to accomplish some of those goals. This is as it should be, but is often overlooked in the planning of and expectations for projects.

HPTP (The Habilitation Personnel Training Project)

This project came about because goals and needs of large numbers of people coincided. Some goals were realized in Governmental legislation and some in University affiliation, but some were not realized but only recognized as needs. It is in stating these needs where goals are most obvious for HPTP.

According to the 1975-76 proposal, "The needs, as identified in the initial project proposal are:

1. A need for more personnel.
2. A need for better trained personnel.
3. A need for differentiation of the type of training required for various habilitation roles.
4. A need for systematic delivery of training." (p. 1)

In the next paragraph the proposal says, "We believe that our training model will provide:

1. Greater capacity to meet increasing demands for Habilitation Service Personnel.
2. Improved training content based on analysis of identical roles and the teaching process.
3. A more comprehensive scope of training emphasis through differentiation of school based roles.
4. A more systematic approach to implementing training which is *generalizable* to other training institutions." (p. 1, italics added)

On p. 14 of the 1972-73 proposal, the following statement appears, "The unique contributions of this prototype model will be the development of systematic, sequential training materials which are *generalizable* to teacher training programs across the nation, and a structure which can be used as a model for program development in the field." (italics added)

In commentary, need number 4 "a need for systematic delivery of training," if met, would satisfy needs 1 and 2 for "more" and "better trained" personnel under the condition that systematic training be efficient and thorough. To some extent systematic training should be efficient and thorough just because it is systematic. Of course, "more" and "better trained" personnel can result simply from the application of training where none existed. This is the extreme case of a zero base rate being established for producing graduates against which any increase in graduations would be an improvement. Obviously, the basic question is, "What was the former rate and was there a change with the systematic training?" This is an intellectual evaluation question.

The social evaluation question is, "What qualitative difference does the training make?" Systematic delivery of training, whether it results in more or less teachers, sounds better than "unsystematic" instruction (which is implied to exist by the very use of the term "systematic"). To be sure, university curricula do not come from one mind nor even very often from one departmental faculty committee concentrating on course development. The responsiblity for courses falls mostly on each one who teaches. Therefore, "systematic" instruction embracing the whole domain ought to be superior to "unsystematic" instruction primarily in its conceptual coherence, its lack of unconsidered redundancy, and its unitary viewpoint. These same characteristics, however, could be its major flaws. The individual instructor's knowledge, motivation, and communicative ability might count for more than "systematic delivery of instruction." In short, by any standard available, variable results may be had depending on the conditions and attributes of the instructive act. The ideal combination would seem to be a systematic delivery system supported by high instructor knowledge, motivation and communication. Do the materials encourage this? An answer to this social evaluation question would be data supporting the qualitative difference made by these materials.

The four claims about what the training model will provide are tied only generally to the four needs identified. Claim number 1, "Capacity...to meet...demands..." relates to need 1, *more personnel*. Claim number 2, "improved training," and number 3, "A more comprehensive scope..." both reply to need 3, "*differentiation...of training*." They also are ambiguous since need 3 is ambiguous. They depend on whether need 3 is interpreted as "differentiation" within broad qualifications of a single job, or across the qualifications set for stable and identifiably different job roles. The scope of claim 2 and 3 is smaller in the first case and larger on the second. To tackle the larger scope problem, considerable field study of the varied roles would seem appropriate. Some field study also seems appropriate for the smaller scope of variations within a role. However, knowledgeable individuals who are experienced in the role could provide considerable insight into that role without field study. Aside from the ambiguity in interpreting need 3, claims 1, 2, and 3 clearly are claims to meeting to meeting all four specified needs.

Finally, the 4th claim, that of "a systematic approach which is generalizable," seems to be a statement that the model meets needs 1, 2, and 4-more and better training, systematically delivered. The unsecured point here is what is meant by "generalizable?" It could mean, "easily usable" by any special education teacher and teacher trainer in the U.S. Other more technical definitions might also have been considered. Usually these have something in common with the "easy to use" definition. These technical definitions require intellectual evaluation data to support the assertion that the training really is generalizable.

This fourth claim is also supposed to meet a need not specified in the list of needs. Specifically, it suggests that not only will this systematic training meet needs within a university, but it will also be "generalizable" to teacher training programs across the nation.

The need for such training throughout the field could be the unifying need recognized by those who are the sources of goals for HPTP. Certainly, those mounting Federal support must have a considerable interest in generalizable training. Further, training materials already developed, provide a significant head start on solving the teacher educator's problems. Therefore, other universities would also have a pre-existing need for generalizable training. Under this viewpoint the term "generalizable" would include the idea "adaptable."

This raises a practical problem for evaluation. To carry out evaluative work, stable and consistent methods and materials must be used. A paradoxical situation arises if generalizable (adaptable) materials and training are purposely developed: the stability and consistency of methods and materials decreases with each adaptation rendering evaluation less possible. In short, generalizable (adaptable) training becomes less evaluatable.

To sum up, the sources of goals for HPTP are varied and sometimes in conflict. This conflict is not by design but rather due to the views of and formulations about what needs are to be met. Such diverse views come about because of the setting in which needs are recognized. It seems clear that the major goals for HPTP can be stated in the sentence:

> "HPTP will produce a generalizable teacher training model with materials which will result in systematic training for more and better teachers of the secondary level handicapped."

Operationalizing this sentence is a major undertaking, only part of which has been accomplished by HPTP. Some of the problems faced by HPTP have been alluded to in this discussion. Some of the problems have been peculiar to HPTP. Despite slippages in meeting goals, some of the major problems have been met and solved by HPTP. The following pages will explicate the way these were accomplished.

ACTIONS TAKEN AND OUTCOMES RESULTING

The actions taken by HPTP have been divided into four different spheres. Figure 1, p. 17, the Project Planning Model, has been used consistently throughout the Project, in one form or another, to guide thinking and action. The Content Model, Figure 2, p. 53, took nearly two years of activity to finalize. The Training Model, Figure 3, p. 57 has taken the remainder of the Project's life. Concurrently with the training model, the Evaluation Model was developed. Part of it (see Figures 4-6, pp. 103-105) consisted of three phases: internal (within HPTP) appraisal, expert (and external) appraisal, and field testing. Part (see Figure 8, p. 115) consisted of a review of project tasks. Also concurrently during this last year, the Dissemination Model, Figure 7, p. 113, was finalized. This model was based on governmental regulations regarding how projects can disseminate products, on governmental copyright regulations, and on a desire to have colleagues be aware of our work. These four conceptual spheres have adequately guided planning and execution throughout the life of HPTP.

The following pages include a narrative history of the Project with evaluative commentary. This provides an overview and a framework in which to view subsequent sections. Following the Planning Model's schema, there are four sections which treat each model. There is a short section on the Content Model. The complexity involved in this model makes it necessary to refer the reader to the Final Planning Report for details. There is a discussion of the Training Model. Some in the Project saw this as the central goal for evaluation in the Project. A large section relating to the Training Model data discusses the numeric data collected under the Evaluation Model. Finally, there is a review of activities carried out in regard to the Dissemination Model. These discussions are preparatory to a summary discussion and recommendation section which completes the report.

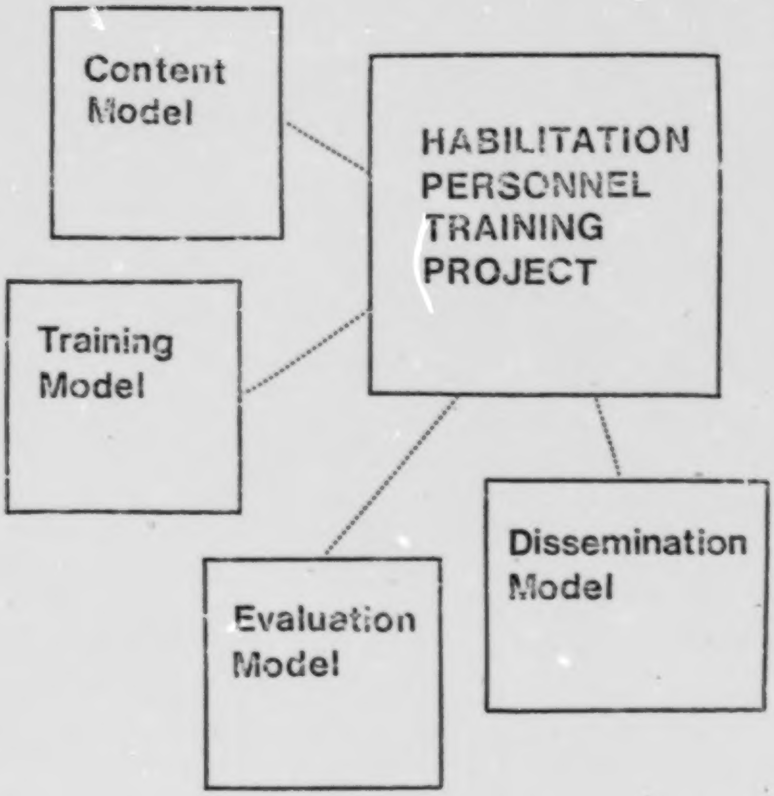

Content Model
HABILITATION PERSONNEL TRAINING PROJECT
Training Model
Dissemination Model
Evaluation Model

FIGURE 1: THE PROJECT'S PLANNING MODEL

An Evaluative History of the Habilitation Personnel Training Project (1971-76)

James P. Girard

The intent of this narrative is not only to give the reader an integrated view of the events and decisions which led to the creation of the HPTP's modular training program, but also, it is hoped, to provide some insight into the sort of trail-and-error evolution which has to be part of all mortal struggles such as this---the mistakes that have to be made in order to see which way not to go, the unforeseen outcomes of those mistakes, and the ways in which they can be dealt with. It would be nice to think that those embarking on such a project would be able to benefit from this narrative by avoiding the same mistakes. Our experience, however, is that this is unlikely. In any event, what is perhaps more important than avoiding mistakes is avoiding the arrogance of believing that one will make no mistakes, or---worse---of failing to recognize when one has made them, or---worst of all---failing to learn from them and to proceed with one's original goals.

The observations and evaluations offered here are those of the Project's editor, who was the longest-employed full-time staff member during the life of the Project. It may also be relevant to note that this editor came to the Project with no credentials or background in special education and---being blessed with a Director who was more concerned with the quality of production than with whose name appeared on which piece of paper---was able to operate in a professional, task-oriented manner, standing somewhat apart from the push-and-pull of academic and departmental hierarchies. It has been necessary occasionally in this narrative to mention individuals by name, for the sake of clarity. Judgements of decisions made or actions taken by particular individuals are meant in no way to reflect on the competency or dedication of the individual in question. It is simply the case that some decisions and actions---through the advantage of hindsight---can now be seen to have been erroneous or even occasionally wise. It is the case that every individual associated with the Project---including the editor---was sometimes right and sometimes wrong, as it turned out. As noted above, mistakes are virtually inevitable. What is more important---and what this narrative will attempt to focus on---is the outcome of the mistake and what is eventually done about it.

The Origin of the Project and the Birth of the Lattice

The Project actually began as early as 1970 when Professor Gary M. Clark and one of his doctoral students, Warren Schoonmaker, put together a very small planning grant to design a prototype training program focusing on "new professional roles" in secondary special education----

especially high school work placement personnel. This was a response to the lack of full-scale training programs, at that time, for secondary special educators, especially those involved in work-study programs. The initial proposal stated, in fact, that "no attention will be given during this planning period for the training needs of the classroom teacher in the high school program for the retarded."

Besides the training program itself, they proposed (1) to develop a recruitment system for trainees (full-time only), (2) to consult with trainers and trainees about training needs, and (3) to provide for program evaluation. The heart of the program, however, was to be a graduate-level training program composed of a core sequence of "two or three courses," plus some training in related areas and, finally, a practicum or internship. Although some "exploration of nontraditional delivery" was proposed, the idea was basically for a traditional graduate school curriculum focusing on the areas of need. The core courses were to deal with the areas of adult adjustment of the handicapped, rationales and instructional approaches for curriculum development (with emphasis on prevocational curriculum), problems and techniques of work-study, and parent counseling. The "related areas" included student characteristics and rehabilitation, human development, diagnosis and evaluation, job analysis, community relations and behavior modification for prevocational training. Clark and Schoonmaker, at this point, intended to develop the core courses as essentially normal graduate-level college courses, with some"programmed instructional materials for selected content areas basic to work placement."

Given the limited resources and staffing requested, this seems a reasonably ambitious proposal---and the need for the proposed outcome was evident (and still is, to some degree.) During the year 1971, however, things were happening in Washington. For one thing, "career education," very broadly interpreted, became the focus of the Office of Education, while the general problem of providing secondary-level habilitation services became a target area of the Bureau of Education for the Handicapped. In addition, Clark and Schoonmaker, during the initial planning, had run up against some of the realities of certification, funding, and so on, which would influence the usefulness of any proposed training program as specialized as the one they had originally conceived. Also, they had been exposed (through James F. Budde) to the possibilities of systems analysis. The result of all this was that, in December, 1971, a much more ambitious proposal was submitted for a two-year project to design, develop, implement and evaluate a systematic but generalized prototype program for training secondary special educators. The trainee population was now to include both pre-service and in-service students, was to be generic to all handicaps, and was to prepare trainees for the work-study or classroom teacher role (or for some hybrid roles).

The planning done during 1971 had involved the use of a "lattice system,"---a kind of elaborate, graphic outline format, in which complex operations are broken down into more and more specific operations or behaviors by means of boxes connected by arrows. The bottom row of boxes-called baseline cells-represent the most specific operations. It is a good system for showing graphically the interrelationships of various specific operations involved in some broad process---in this case, the habilitation process, which was the generic term used to describe what the teacher and the work-study coordinator do. As a training tool, a lattice lends itself most effectively to a practicum format, in that it shows the specific steps the trainee takes on the job and during the training experience---and this is, in fact, what the Project planners had in mind. The plan was to develop a set of "instructional competency units" (ICUs), each of which would correspond to one of the base line cells in the lattice. The ICUs, at this point, were described as "study guides," which outlined a training curriculum to be delivered with an emphasis on practicum -type experiences, plus some traditional coursework and with some package-type materials filling in here and there, wherever appropriate. The ICUs, then, were seen essentially as elements of a curriculum guide, not as actual training materials. This is an important distinction, because a shift in definition of the term "ICU" occurred during 1972, which affected every other aspect of the Project, resulting in a radical shift of Project goals, and adding at least another two years to the Project's scope.

Problem Number One: The Uncontrolled Evolution of Concepts

Much of 1972 was spent in consultation with other program developers in an effort to arrive at a workable format for the ICUs and the training materials themselves. At that point in time, it was inevitable that such consultation would involve a great deal of exposure to advocates of competency-based teacher education. Furthermore, most of the materials encountered (such as those developed at the University of Houston) had a heavily behavioral, heavily programmed design. Deceptively simple, this was a design that was initially attractive to a Project which consisted of a single professor and by now, a handful of doctoral students, plus a couple of master's students working on stipend---none of them on a full-time basis. It had already been proposed to "select priority items" from the developing model (lattice), draft ICUs for those items and do initial pilot testing. At some point, it was decided to develop a competency-based modular training package, following the Houston design, on the topic "Writing Behavioral Objectives." The resulting package was referred to as an ICU.

It is important to note that this differs from the original conception, in which an ICU was defined as a fairly broad element in an overall curriculum. ICUs originally were to include such specific training materials as a package on writing behavioral objectives. As a result of the experiment with Houston-based materials, however, Project planners began gradually to think of ICUs (the basic conceptual units of the lattice---the base-line cells) as much more specific training objectives, each of which could be developed into a modular format. This in turn changed the conception of the Project's final product from a curriculum guide incorporating some training materials, to a full-scale, self-contained training program in modular form.

There were other unforeseen outcomes of this decision, which would affect Project working procedures for much of the Project's history. For one thing, it dramatically increased the projected size of the lattice. As noted above, Project planners and conceived of the ICUs, or baseline cells, as representing fairly broad portions of an overall curriculum guide--- with each baseline cell roughly equivalent to a single three-hour course, in traditional terms. Now, however, if "Writing Behavioral Objectives" was itself an ICU, then the baseline cells were to refer to much more specific levels of behavior, and each was to be much less than a traditional college course in scope. By January of 1973, in fact, Project workers would be routinely talking in terms of 300 ICUs themselves. It is important to remember that work had begun on the initial modular package---"Writing Behavioral

Objectives"---on the assumption that this would represent only a portion of a somewhat larger course. Thus, it could be seen as standing on its own---a supplemental unit to be "plugged into" a larger course curriculum. As this package began to be regarded as an entire ICU, however, and as the eventual training program came to be conceptualized as an integrated set of such packages, the nature of the individual packages changed. Instead of a discrete item which could be designed without reference to any of the others, each ICU came to be seen as one element in an integrated sequence. It seems clear that persons designing such packages would need to know exactly where each was to fit in such a sequence, and what the interrelationships would be, in order to avoid overlap, gaps or redundancy. In fact, however, because the lattice was now seen as largely incomplete, nothing was really known about the eventual overall structure. This aspect of uncertainty about the true scope of each ICU would stymie Project writers for most of the next two years.

That the shift in definition of the term ICU was allowed to take place at all is partly attributable to the nature of the first package developed. The topic of "Writing Behavioral Objectives" is one of the easiest to deal with in a modular (or even programmed) format. It is probably the topic which has been worked over the most in such formats, since it is often the one chosen by those beginning to design modular materials, as a way of "getting their feet wet." Because it lends itself so easily to this format, it tends to reinforce, for novices, the initial attraction of the format as well as giving the erroneous impression that other topics may be dealt with just as easily. Frankly, advocates of competency-based education usually do not go out of their way to dispel this impression. In any event, by the time they had completed the "Writing Behavioral Objectives" ICU, it is not surprising that Project planners were thinking (perhaps hopefully) that it would serve as a model for development of all future ICUs. And, in fact, despite many technical changes in format, this basic ICU design remained a powerful influence on Project product development until the fall of 1974, when it was consciously set aside.

One outcome that persisted into the Project's final year was the peculiar twist the modular design gave to project evaluation. The term "evaluation" was one which was never adequately pinned down during most of the Project's life, partly because it appeared to have more than one meaning. During the period when planners were consulting with those in the field, a heavy emphasis on "program evaluation" was uniformly recommended. However, it was never really made clear whether this meant evaluation of the Project"s working procedures and progress objectives, evaluation of the Project's final product (fieldtesting of the training program), or evaluation of the work of the trainee going through the program after having been trained. All these possibilities came into play at various times and the word "evaluation" was used generally to refer to all of them, somewhat indiscriminately. The Project's *First Formative Evaluation Report*, for example, dealt with fieldtest data on the ICUs while the *evaluation section* of the lattice concerned skills needed by the trainee to evaluate the work of handicapped students. Because fieldtesting data were based on the performance of trainees within the ICUs, evaluation of *trainee performance* inevitably became confused with evaluation of the materials themselves. Unfortunately, the apparent ease of writing tests for the"Writing Behavioral Objectives" ICU gave Project planners false expectations about the ease with which test components could be devised for future ICUs, and, in any event, because of the confusion of terms, the whole question of program evaluation came to center around the writing of tests for the ICUs---an area in which the Project as a whole possessed little experience or expertise. Periodic attempts to bring in consultants in this area failed because of the confusion about what was intended; the consultants were never entirely sure what they were supposed to *do*.

Problem Number Two: The Failure to Control Communication and Decision Making

There is no question but that the early rapid growth of the Project---thanks again to the shift from a curriculum guide concept to the idea of a full-scale modular training program---contributed greatly to the confusion of concepts and terminology. When the Project had consisted only of Clark and Schoonmaker, it was possible for the two of them to have enough day-to-day contact to maintain the necessary agreement about the concepts and terms they were using. As the scope of the Project grew, however, many additional people were brought in, including three additional doctoral students (Stuart Schwartz, Patricia Louise Sitlington and Sherman Yacher), a number of present and former master's students who served as ICU writers, and, eventually, a full-time editor. Most of these new staff members were never fully informed about old conceptual agreements and had, therefore, to work things out from context as well as they could. Gaps began to appear.

A major part of the problem---and one which was recognized very early, but turned out to be difficult to deal with---was the superimposition of departmental and academic ranking on the Project's arrangement of staff. When Clark and Schoonmaker had been alone, they had been able to work as a senior and junior colleague, working out conceptual matters informally. The addition of three new doctoral students, however, created a vastly different situation, since each of them expected an equal part in decision-making.

Clark recognized very early, as did most of the others, that there was a need for a full-time director or coordinator, to make the necessary decisions, but funds for such a position were not immediately forthcoming, so some other temporary answer had to be found. Doctoral students were given administrative author' y in several organizational areas. When the simple division of authoity among these administrative staff failed to work, the Project eventually turned to elaborate work-flow systems with muitiple checkpoints.

With the Project becoming bogged down at the conceptual and decision-making level, it should not be surprising that the Project's other new employees---those hired to write and edit the ICUs---were left to work in a kind of conceptual limbo. Their problem was oddly complicated by the fact that these first Project writers were nearly all present or former master's level students in Special Education (the lone exception being a journalism major who nevertheless was hired, because she had experience working in a sheltered workshop.)

Along with this, there was a feeling among the administrative staff that the writers needn't be concerned about the problems of the lattice but only needed to write what they were told to write, once it had been determined, because, as master's students, writers were expected to know enough to be able to do their writing with minimal guidance.

Consequently, the development of training objectives for modules was left to the writers who, in fact, were virtually ignorant of the lattice where these objectives were to be spelled out in logical fashion. Unfortunately, the editor at this time made the mistake of assuming that all these people with backgrounds in special education were in agreement about the terms they used, and so failed to take standard editorial steps to have the lattice staff define terms at this juncture (a failure which contributed to the Project's problems for one year). Because the only materials development in progress was that being done by the writers on their ICUs, while Clark and the doctoral students worked primarily with the lattice, the editor also was led to believe that ICUs being designed by the writers constituted the sum total of materials required by the Project. The original idea of a series of curriculum guides for an extensive practicum, supported to some degree by traditional courses and packaged materials, had now been virtually abandoned; the packaged materials had taken over to the exclusion of everything else.

One outcome of this sort of staff fragmentation was that the writing staff, who knew least about the basic concepts underlying the lattice, began by default to assume all actual decision-making power with regard to ICU design. At the same time, finished ICU materials were routinely subjected to a "content review session," at which all the lattice staff, in seminar fashion, took the ICU apart piece by piece, rarely putting it back together again. It quickly became clear that any ICU which had its own internal logic and consistency (whether or not it bore any relationship to the lattice) fared much better at these sessions than did ICUs which attempted to reconcile the differing views of the reviewers. Thus, writers began virually to ignore the arguments of the lattice staff and instead produced modular-type materials on the assigned topics, in formats with which they were familiar from their own training experiences. In general, the emphasis in ICU design shifted away from the problem of establishing training objectives to the somewhat simpler matter of finding or designing training activities (with the objectives written afterwards, to fit the activities.) Because this led to ICUs which were more interesting and coherent, in their own terms, the lattice staff---unsure of its own ground---had difficulty rejecting such efforts.

Inevitably, many of the decisions made in this way by the writing staff were at odds with the Project's original goals and concepts, and yet---because some of the ICU materials appeared to make more sense, in some ways, than the lattice itself---these decisions began feeding back into the lattice and distorting it.

An example of a serious error made at this time because of this situation involved the selection and structuring of objectives within the ICUs. Under the original practicum-based conception of the training model, objectives were to call for the trainee to actually perform various steps in the teaching or habilitation process. Thus, for example, the trainee would demonstrate his or her ability to design behavioral objectives for the handicapped students. In order to reach this point, it would often be necessary to provide background information or activities (e.g., articles on how to write behavioral objectives) but the trainee would actually be tested on his or her ability to do the task in question, not merely know how to do it. As noted above, this scheme worked well with the "Writing Behavioral Objectives" ICU, but in later ICUs, where the target behaviors were less specific and less easy to sample (e.g., "Selecting Instructional Methods."), the amount of background information which might be relevant to completion of the final objective multiplied rapidly and no single item of background information could be accurately predicted to enable successful completion of the objective. In other words, while one or two specific readings on writing objectives might directly prepare the trainee to write objectives, no similar materials could be found or created to prepare the trainee for selecting the appropriate teaching method for a given situation. Instead, it seemed necessary for the trainee first to become generally familiar with the qualities and characteristics of as many different methods as possible, in order to be able to make a meaningful choice. At this time, as noted, Project writers were focusing most of their attention not on the objectives themselves, but on activities. They were under some pressure to produce materials which could be easily tested, based on the "Writing Behavioral Objectives" model. Needless to say, it is far simpler to find activities for and to test objectives dealing with the trainee's knowledge of teaching methods than it is to design and test objectives calling for the trainee to do the selecting. After the door was thus opened to substituting "knowing" objectives for "doing" objectives within the ICUs, such objectives began to creep into the lattice as well.

The shift from a practicum design to a modular design aggravated the confusion. If baseline cells of the lattice represented performance steps to be accomplished by the trainee in a "doing" situation, then the lattice spelled out not only the content to be trained, but also presumably the training sequence. The modular format, however, with its now primarily "knowing" objectives, implied an entirely different sort of training model, involving far more preliminary academic training before the practicum. It would become necessary to transform the existing content model (the lattice) into this new training model (represented by the new ICUs) and to integrate the two, but at this point, the lattice was instead distorted to represent the demands of both models. It was further distorted by a basic confusion about the domain it was intedned to deal with. Originally, it was to depict what the secondary special educator does in teaching a class or conducting a work-study program. No provision was made, however, for defining in systematic form such things as what the handicapped student was to learn, what the trainee was to be taught (as opposed to what he or she was to do, after having been taught), or what the instructor using the ICUs was to do in the course of training the trainee. Gradually, at points where questions were raised about such matters, items belonging to these domains also began to creep into the lattice, compounding the confusion.

Judging from memos from this period, it seems clear that Schoonmaker was aware of the need for integration of diverging Project components, and also of the need to think of the lattice in multidimensional terms, rather than straight-line logic, but, partly because of the conflicts among the lattice staff and partly because of his own discursive style, he was not successful in communicating these concerns to staff members. Instead, most Project staff communications were over terminology with consensus eluding the staff's efforts. At the same time, memos show that inadequate funding for production (since the need for production had outstripped the Project's original goals) left four half-time writers sharing a single typewriter. On top of that, the Project's only half-time secretary struggled to keep up with the many memos, the review materials required by the lattice staff, and the typing of final drafts of the ICUs themselves. This last task lagged by two to three months, and distribution of sufficient copies for fieldtesting was inevitably an emergency situation, with all staff members staying late to collate and staple. Even so, far more resources were being devoted to actual production than to fieldtesting, which was being largely handled by Clark himself, in his "spare" time.

A Solution That Failed: The Implementation Model

That the Project was virtually at a standstill was evident to everyone in the Project by the fall of 1973. The continuation proposal, submitted in October, 1973, was by far the heftiest document ever produced by HPTP, but it was heavy on graphics (primarily loading charts and lattices) and light on narrative. What narrative there was dealt with what the Project *intended* to do, plus rationales for why it hadn't yet *done* it. It was clearly a plea for greater funding, with the intention that more money would answer the Project's problems. Perhaps, it is only by hindsight that fundamental, conceptual, and organizational difficulties seem to glare from every page. In spite of these, and as requested, funding was increased more dramatically than at any other time in the Project's history. (Ironically, a year later, when the Project actually had reached firm ground again and submitted a report which clearly indicated that substantial, significant work was underway, funding was cut so severely that the writing staff had to be reduced by half.)

On the other hand, the continuation proposal provided evidence that Project planners were aware of many of their problems, and it detailed a complex, systematic approach which had been designed to deal with them---the Implementation Model.

The Implementation Model arose partly out of a tendency for the Director and lattice staff to view the Project's problems in terms of low productivity on the part of the writing staff. The recognized need for a full-time coordinator, for example, tended to be described as a need for more oversight of the writing process. Schoonmaker, who had been most deeply involved in the lattice from the beginning, was perhaps most aware that there were serious problems in that quarter as well, but nevertheless, it was he who set the Implementation Model rolling.

His original plan was to apply Budde's working draft of an accountability system then known as FACTS (Formative Assessment and Continuous Tracking System) primarily as a research tool, to find out exactly where time was being lost. FACTS called for each worker to describe, in terms of activities, objectives and products, exactly what he or she did, to designate these various activities and products by number, and then to chart exactly how much time was spent on each. A new administrative staff member, Greg Weisenstein, was given the assignment of collecting these detailed work descriptions and putting them in some kind of order, so as to provide a comprehensive, graphic representation of what workers in the Project were doing.

Weisenstein thus unfairly came to be regarded as a latter-day Dr. Frankenstein, because he was unfortunate to be one of those new Project employees who was assigned complex duties without adequate preparation. Not being an expert on systems or graphics, Weisenstein chose to represent the whole thing as one long flow-chart, showing the movement of each ICU, from identification of content to fieldtesting. The result forced an impression of an inflexible sequence of steps to be followed in the production of every ICU. Although Schoonmaker distributed a memo stressing that Weisenstein's charts were meant to depict Project procedures, not to dictate them, the staff for the most part began treating the massive flow-chart as if it were, indeed, a detailed directive for what everyone and especially every writer, was supposed to be doing. This was complicated by the fact that some portions of the Implementation Model didn't actually describe what had really been going on! The persons filling in that portion either didn't know what procedures were being used or they described what they felt ought to be going on. In addition, various administrative and lattice staff members took pains from the outset to make sure that the model spelled out not only procedural steps but also how authority was divided. In many areas authority actually overlapped. Confusion over authority and over content in the ICUs resulted. Not surprisingly, production ground nearly to a complete halt. The accountability charts being filled out by the writers revealed that they were spending most of their time either in review meetings or in trying to set up such meetings, and very little on actual writing. Memos indicate other staff members were struggling to organize their thinking about the project. No ICUs were reaching completion.

The major practical outcome of the Implementation Model was to firmly divest the writing staff of the de facto decision-making power they had been exercising---but without replacing it with anything. Thus, for nearly a year---from fall, 1973, to the spring of 1974---virtually little progress was made at all in delineating a workable training model.

The first break in this static situation occurred in December of 1973, with what came to be known as the "writers' revolt." Officially, the writers were advocating a restructuring of the Implementation Model, and, in this spirit, Weisentein was assigned to redesign the Model so as to take the writers' perspective into account. In fact, however, the meeting was the beginning of the end for the Implementation Model, since it was apparent to everyone that the writers' problems were only symptoms of broad underlying difficulties, of which the Implementation Model was the most apparent part. Again and again, those in the Project had come up against the same problems: the unfinished lattice which prevented any sense of the ICUs fitting into a coherent, integral whole; the failure to resolve conceptual and terminological gaps, which prevented both the lattice staff and the writing staff from having confidence in the value of their respective products; and the lack of a full-time coordinating authority, which prevented either of the other problems from being adequately dealt with. Because these problems had been around for a long time, the writing staff was not optimistic about the prospects for real change. The writers underestimated the rest of the staff.

The Solution(s) That Worked

Although oversimplified, it would be fair to describe the distribution of administrative authority within the Project at this point in terms of a triad. Each of these three "power nodes" initiated activities during the spring of 1974, which, taken together, would effectively solve the most serious of the Project's difficulties. Ironically, these activities were all undertaken without consultation between the two junior members of the triad while informing and getting approval from Clark, the senior member. As it turned out, the various activities fit together as effectively as if they had been fully coordinated.

First, Sitlington and Schwartz, acting on the belief that the lattice had to be completed in order to provide a solid conceptual basis for the ICUs, set up a two-day workshop, at which all staff members would devote full time to an intensive effort to complete the lattice once and for all. Each staff member was assigned to one of four groups working on the four different lattice sections, and all other work was suspended for the duration.

Secondly, and at the same time, Schoonmaker, responding to a need to begin integrating diverse conceptual components in line with the implicit logic of the modular format, assigned the editor to produce a module on the "Referral Process," combining a number of ICUs which were currently in progress. ("Identifying Referral Procedures," "Implementing Referral Procedures," "Identifying Referral Resources," and so on.) Schoonmaker understood that the ICU topics, tied in as they were to the baseline cells of the lattice, were too limited to make real sense either to the people writing them or to the people using them. (Fieldtesters, in fact, had already begun to demand integrated sets of ICUs). The training model implied by the modular design had its own logic, different from the lattice, and Schoonmaker felt it would be easier for writers, reviewers and fieldtesters if the materials being produced reflected that logic, instead of being tied to the baseline cells. Combining several pedagogically related ICUs in one module also resulted in savings of research and writing time, as well as a more efficient format. The "Referral Process" module was finished just before the lattice workshop was to be held, but Schoonmaker elected to delay releasing it to the rest of the staff until after the workshop, so as to avoid confusing those still involved with the lattice.

As it turned out, however, the new modular design indirectly had a revolutionary effect on the design of the lattice. The actual content of the "Evaluation" section had already been fairly well established. That is, the necessary steps and procedures (collecting and interpreting data on student performance, and so on) were already known---all that remained was to complete the lattice depiction of how these steps fit together in a logical pattern. This group decided to try working at this in reverse---that is, by asking what modules would probably need to be created to cover the content of this section of the lattice, then afterwards representing those modules in a lattice format. In effect, this skipped the step of designing the content model in final form and moved directly to planning the training model (recalling that a training model was intended to be the Project's primary goal in the first place.) This approach worked even more easily than had been anticipated, and the "Evaluation" section was the first to be completed at the workshop.

Following the workshop, Schoonmaker and Girard sought to consolidate what they felt was a viable new direction for the Project. The "Referral Process" module was presented to the staff as a new working model to replace the old model ICU, "Writing Behavioral Objectives." After collecting lists of all the baseline cells identified at the workshop, where as Schwartz and Sitlington had hoped, the groups had done substantial work in advancing completion of their sections under the old logic, the editor met with the administrative and lattice staff to negotiate a tentative list of modules, which would be similar in organization to the "Referral Process," and which would reflect the content---but not mirror the organization---of the baseline cells. Because of some remaining areas of confusion---primarily relating to testing and to the preponderance of "knowing" rather than "doing" Objectives---agreements were highly tentative at some points. The chief effect of these meetings, however, was to emphasize the need for integration of the two models, content and training, particularly in the transformation of content into a workable training sequence, as well as highlighting the multidimensional nature of the lattice.

These activities were supported by Clark, who circulated a memo establishing these new priorities for the Project: (a) to finish the still-unfinished "gray areas" of the lattice; (b) to concentrate on transforming lattice content into the modular training logic. Clark also took the third major step which contributed to the overall solution, by making it known that Keith Edmonds, one of the writers, would become the Project's first full-time coordinator as of June, 1974.

Edmonds had been working as a writer since the summer before. His presence within the Project, however, was the result of an experiment advocated by Girard. Previously, all Project writers had been persons with some background in special education. The Project's official policy---virutally identical to that of most such operations within the academic community---was to hire people with credentials in the field and then "teach them to write what we want." When a full-time writing position came open in the summer of 1973, at a time when the job market produced a large number of qualified applicants, the editor, pointing out that "writing is what we are least prepared to teach anybody," urged that someone with obvious organizational and writing skills be hired, and then taught the necessary content or provided the necessary resources. Edmonds, a master of arts in teaching who had administered the Harvard Africa program the year before in Kenya, was selected. In promoting him to coordinator, Clark was making a conscious decision to employ someone with experience and skills dirictly relevant to the position. This was a policy which would be followed until the end of the Project, whether it was a coordinator or a typist who was to be hired.

Clark also arranged to have Dr. Mel Semmel of the Indiana Center for Innovative Teaching of the Handicapped visit the Project in May, 1974. Semmel spent an eventful two days with the Project, forcing staff members to explain and defend virtually every aspect of their work and planning. At the end, he stressed the need for the Project to make some of the decisions it had been holding in abeyance since very early in its history, and to make some of those which were necessitated by the conceptual shifts which had occurred. Such decisions included a closer definition of the Project's target trainee population, decisions as to the role to be served by instructors using the Project's modules, and decisions affecting the use of tests within the modules. Many of these decisions simply reflected a need to redefine Project goals in terms of conceptual changes which had occurred since the original grant proposal. At the same time, Semmel had surprisingly high praise for the quality of work done so far, and especially for the performance of the writing staff, whom he elected to meet with privately, after which he recommended far more autonomy for the writers in decision-making areas, with other staff to serve largely as content resources and Clark or Edmonds to provide some systematic way of ratifying and formalizing production decisions as they were made.

Perhaps surprisingly, there was no opposition to Semmel's recommendation of writer autonomy, nor to Edmond's promotion. What had too often been viewed as a struggle between the "good guys" and the "bad guys" for control of the Project, in fact, turned out in retrospect to be a concerted effort by everyone to solve the Project's problems. Once the lattice was essentially complete, once the modular format was understood and properly integrated and once a full-time coordinator had been established, the administrative and lattice staff freely relinquished control of the production process to those with the appropriate skills, while they (with the exception of Yacher, who joined the writing staff) turned their attention to the demands of fieldtesting. This willingness to do what was necessary to accomplish the Project's goals, instead of clinging to the kind of hierarchical power structure institutionalized by the Implementation Model, was an important characteristic of the Project, particularly in its last two years, and contributed greatly to the quality of the Project's work and products.

The Last Two Years: A Few Remaining Problems

The summer and fall of 1974, was a time of change no less significant, in different ways, than those described above. As already noted, turnover within the writing staff resulted in persons with special education content background being replaced by persons with substantial writing and organizational skills. Schwartz and Schoonmaker both left the Project, and Marianne Notley, a full-time media specialist, was hired. Perhaps most significantly, the writing staff managed to secure working quarters in a different building, on the other side of the campus from the special education department. This produced some practical advantages---such as removing Project secretaries from the departmental pool at a time when production of materials was increasing dramatically---and a major psychological advantage in removing the writers from the hectic atmosphere of an academic department, enabling them to be completely task-oriented. The physical separation had the effect of making tangible

Semmel's recommendations of writer autonomy, and increased morale dramatically. Edmonds accompanied the writers in the move, while Clark and the content resource staff remained behind in their special education quarters (though office space was provided at the new department).

In this new setting, the writers adopted the team approach to module identification which had been used at the lattice workshop, and each section of the lattice was gone over carefully and translated into modules, which were assigned to pairs of writers, so that virtually all work was undertaken in a team context. Master's students assigned to the Project during this period were used not as writers (though some of them did some writing), but as resources, and most of them made significant contributions to the eventual program. The continuation proposal submitted in late 1974, stressed not what was intended, as had the monstrous document a year earlier, but described instead what was being done and how. Contrasted with the 1973 report, it demonstrated a strong sense of direction and self-confidence on the part of the Project workers, as well as embodied a good deal of self-critical historical perspective. As noted above, the response from Washington was to cut funds drastically from the proposed budget, so that the Project's final year (mid-1975 to mid-1976) was accomplished with a half-time writing staff.

The Project's penultimate year was so given over to accelerated module production---which contrasted so favorably with the inaction of the preceding year---that some of the problems which persisted were rather overlooked. It was not until Edmonds left the Project in the summer of 1975, in fact, and a new coordinator was hired, that some of these resurfaced---some of them unfortunately too late in the Project's life to be genuinely solved.

The new coordinator, Alexander Wessitsh, was hired for his skills in the area of evaluation, primarily because it was felt that the production process was running smoothly and that more direction was needed in the area of fieldtesting. Wessitsh's skills, however, enabled him to see that the old confusion within the Project about the diverse meanings of "evaluation" still existed, and that the sort of "doing" objectives originally seen as the content of the modules had been virtually swamped by "knowing" objectives. Although it was too late to completely redo the modules which were finished, the overall program was revised to distinguish more clearly between those modules which dealt entirely with cognitive objectives, and those which actually attacked the skills needed by prospective teachers. In the latter case, especially---but to some degree in all existing modules---this led to finding that existing modular tests were directed at particular module content and insensitive to the whole domain required by the objectives. This was partly because of the confusion about objectives, partly because of the lack of professional experience or knowledge of the content, and also partly because those who had written the tests---primarily the writers---had little background or expertise in the area of test-writing, especially with regard to the peculiar demands of the competency-based modular design, for secondary special education teacher trainees.

It quickly became apparent that the Project had neither the time nor the resources to develop genuinely appropriate tests for each module. It was tempting, this late in the Project, when great effort was being given to securing publication of and arranging for dissemination of materials, to go ahead and include the tests which had already been written, appropriate or not---the implicit argument being that they were "better than nothing at all." Instead, it was decided to take a more difficult course: to salvage what items could still be used, to write as many new items as possible, as exemplars, and to provide prospective users with a testing guide, written for the most part by Wessitsh. These were put into one volume containing tests which would be appropriate for each objective, advice on how to create more items, and some directions about how to administer and evaluate the tests an instructor used.

This decision meant that it would now be impossible to present the final modular program as "complete," in the sense that any instructor might pick it up and use it without having to use additional materials or test items. On the other hand, it was felt that to attempt to make such a claim, on the basis of test items which now seemed clearly incomplete, would be to call into question the value of the program itself. The Project's policy, as it had been all along, really, was to create a product as near as possible to the ideal, given the limited time and resources available, and not to pretend to do more than was actually possible, nor to have to apologize for what was finally produced.

Conclusions

As noted at the outset, it seems unlikely that any project of this sort will be able to avoid mistakes entirely. On the other hand, it seems possible, on the basis of the HPTP experience, to suggest some general methods by which the negative outcomes of mistakes can be minimized, and the lessons learned turned to advantage. These seem, to this writer, the major lessons of the last five years:

> ---It is far more important to agree about the meanings of the terms being used than to simply use the same terms. For this reason, it is necessary to have some means of monitoring the inevitable evolution which occurs with words, and to insure that everyone---everyone---understands what everyone else means by the terms being used, and how these relate to the concepts on which the Project's primary goals are based.

---It is imperative to have a single decision-making authority. It is equally important, however, that whoever fills this role has the time and desire to make a complete commitment to an administrative role. In general, it is important to define operational roles in terms of _the job to be done_, not to simply allow staff organization to mirror the prevailing institutional hierarchy.

---In general, it is vital to hire staff members on the basis of their competency in the task required. This would seem to be an obvious point. If what you want is someone with knowledge of a particular content area, by all means seek out a Ph.D. in that discipline. But if you want a writer, hire someone who can _write_. The same goes for administrators, media specialists and typists.

---It is necessary to be honest about the degree to which one's resources will allow one to approach one's goals. If a trade-off if necessary, it should not be in the area of quality---but it should be recognized as a trade-off and not disguised as a change in goals. A corollary of this is that it is necessary to recognize when past efforts have not truly contributed to the accomplishment of one's goals, even if this means junking a lot of tangible and nicely mediated products.

---Decisions should always be made in terms of the desired goals, and not on the basis of who will get the credit or the blame.

The Habilitation Personnel Training Project proved to be a personal and professional odyssey for all the staff who became truly identified with its mission and process. The agony and the ecstacy of creating and carrying on the Project are only partially reflected in these pages. It is hoped that others can benefit from the reporting of problems experienced and alternative solutions attempted.

The Content Model

Figure 2 is a schematic overview of the content contained in the modules. The format of Figure 2, as well as of its elaborations, is called a lattice. It defines on its baseline the actual activities required to accomplish the superordinate states defined on the ridgeline. The elaboration and subspecification of baseline activities,, and consequently of superordinate states, is given in detail in the Final Planning Report.

The reasons for the content in the lattice are also given in the Final Planning Report. On page 39 of the Report, a summary of goals for the lattice is presented: "...the emphasis of our training programs should be in the functions performed by habilitation personnel rather than on preparing trainees to carry out specific roles" (Schoonmaker and Girard, 1975). Such an emphasis was seen as requiring habilitation personnel to consider needs of handicapped individuals as foremost. On page 67, another reason for the lattice is offered: "...as the lattice becomes complete, ...it is possible to see how the identified competencies distribute themselves in modules...for delivery. Since this is what the lattice was designed to accomplish in the first place, its work is then done." (Schoonmaker and Girard, 1975). In short, these goals describe how the project goals should be met by the development and completion of the lattice.

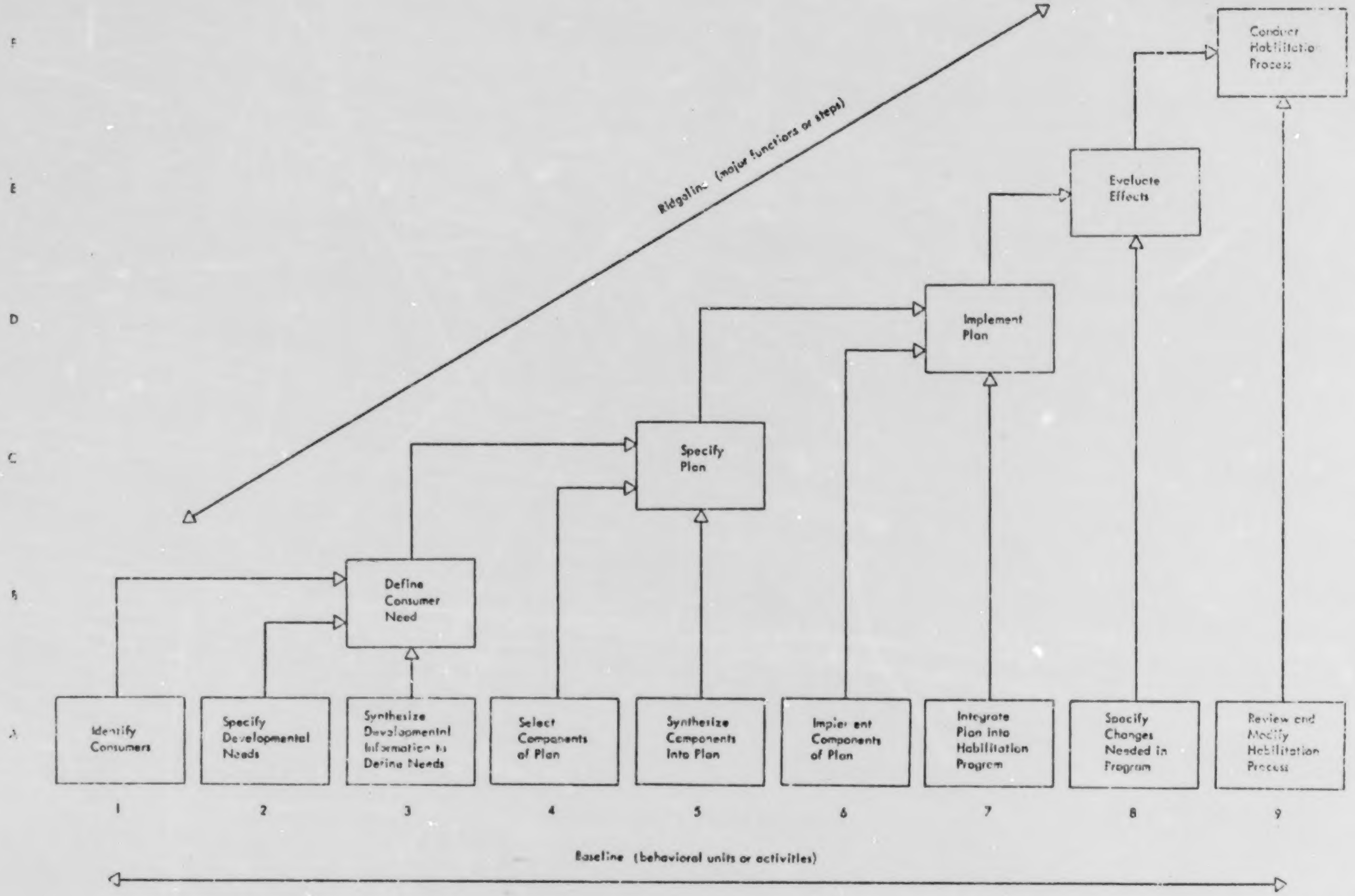

F
E
D
C
B
A
Ridgeline (major functions or steps)
Conduct Habilitation Process
Evaluate Effects
Implement Plan
Specify Plan
Define Consumer Need
Identify Consumers
Specify Developmental Needs
Synthesize Developmental Information to Define Needs
Select Components of Plan
Synthesize Components Into Plan
Implement Components of Plan
Integrate Plan into Habilitation Program
Specify Changes Needed in Program
Review and Modify Habilitation Process
1
2
3
4
5
6
7
8
9
Baseline (behavioral units or activities)

FIGURE 2: THE CONTENT MODEL

Project activity devoted to work on the lattice took about two years. This was no doubt due, in part, to the conceptual complexity of the task, as well as to the difficulty of organizing the task. That the Project has completed the modules is a fact. Did the lattice serve the purposes claimed for it? Dr. Mel Semmel, Project Consultant, talked with Project personnel before the lattice was completed. Nevertheless, he acknowledged to the Project Director ". . . (when questioned) that our project was superior to most other programs that are attempting to develop a competency-based teacher education approach..." (memo dated May 28, 1974, from Gary M. Clark to HPTP Staff.) Similar comments have been offered during the dissemination efforts.

The emphasis on functions rather than roles can be seen in the Training Model, Figure 3, p. 57 . One can also conclude that need 3, discussed early in this report (p. 8) was changed during preparation of the lattice. An outcome of the lattice, obviously, was to shift the focus of upcoming work toward differentiation of training to meet pupils' needs and away form differentiation based on teachers' jobs. Nevertheless, the modules can be organized by role and retain significant role differences.

With respect to the lattice's work being "done," the evidence is that the modules and the lattice are indirectly related. The overall lattice structure shown in Figure 2 does appear in Figure 3 as index letters for the training moduels, "Define Consumer Needs" becomes "N" for "needs," "Specify Plan" becomes "P" for "planning" and so on. Furthermore, as late as the fall of 1975, discussions about revisions of old modules and development of remaining modules contained references to "lattice sections." Thus, for those familiar with the lattice, it provided conceptual foundation, albeit transformed to meet the requirements of the Training Model. Furthermore, the Training Model retains the philosophy of the lattice. In short, the value of the lattice was as a conceptual tool rather than as a diagram for building modules.

The Training Model

Figure 3 is a listing of the modules in the recommended training sequence. As such, it indexes the expected needs of trainers and trainees in preparing for habilitation functions.

As noted in the introductory section to this report, there are several levels of objectives possible in a project where the ultimate target population is students and the penultimate target population (teacher trainees) is the one to which material is directed. In particular, a review of the modules reveals that the training sequence, and therefore the model, has few objectives specified for trainers. Furthermore, Project discussions about the role or function of trainers precipitated comments that the Project could not or should not become involved with the psychology of instruction. As noted in the history and confirmed by review of instructor comments, instructors frequently used modules in manners to suit their needs and situations. To accommodate, more breadth was finally built into module activities as a response to such usage so that trainers might more appropriately incorporate the modules into the form of instruction which suited them.

As completed, the training sequence aimed at the trainees includes background material in the first modules, training activities to sharpen "personal skills" roughly in the middle ones, and "rehearsal"

FIGURE 3: THE TRAINING MODEL

PERSPECTIVES
- N I: The Needs of Exceptional Youth and Adults
- P I: Available Roles in Secondary Special Education
- P II: Career Education
- E I: Program Evaluation

COMPETENCIES FOR INDEPENDENT LIVING
- N II: Adult Adjustment
- N III: The World of Work
- N IV: Home and Self
- N V: Community and Society

NEEDS AND OBJECTIVES
- N VI: Behavioral Domains and Instructional Objectives
- N VII: Assessment Resources and Techniques
- N VIII: Assessment and Interpretation
- N VIIIa: (Pre)Vocational Assessment for the Handicapped
- N IX: Task Analysis

PARAMETERS OF PLANNING
- P III: The Classroom Teacher
- P IV: The Work/Study Coordinator

FACTORS IN PLANNING: RESOURCES
- P V: Audiovisual Equipment
- P VI: Physical Environment and Classroom Design
- P VII: Instructional Resource Personnel
- P VIII: Teacher Aides and Paraprofessionals
- P IX: Referrals
- P X: Content-Related Resources

FACTORS IN PLANNING: PROCEDURES
- P XI: Instructional Methods and Teaching Aids
- P XII: Behavior Management
- P XIII: Communication and Counseling Procedures
- E II: Evaluation of Student Performance
- E III: Evaluation of Program Design Components

PROGRAM IMPLEMENTATION
- N X: Student Needs Assessment
- P XIV: Program Design
- E IV: Evaluation
- I I: Practicum

or simulation activities in the last ones. Practicum caps off the sequence. This trend in the overall sequence towards experiential training was not apparent to field testers, since modules were not used in complete sequence. A number of comments from field test instructors indicated that their goals for trainees included more experience oriented activities in every module. (See the section on instructor module evaluations, p. 60ff, for details). In summary, the instructional psychology which actually was incorporated in the Training sequence was not clear to trainers because they saw only elements of the psychology in the modules they used.

The following pages present the analyses of other data gained from field testing a subset of the modules. There are two reasons only a portion of the modules were field tested. First and primarily, only a subset were complete in time for field testing. Experience showed that a two to three month lead time was necessary for the Project to prepare and for field test instructors to review modules already completed in order to field test them. This period would include initial contact about the module to the final point of an instructor's handing out the modules. Thus modules in preparation could not be field tested rapidly. The second reason for a subset being tested, in any event, was that there were more modules, even with only a portion complete, than there were field test sites and time available to carry out blanket tests. The data, for these reasons, are less than comprehensive.

To conclude, the training model was finished late in the Project. Whether, as a whole, it meets intellectual or social evaluation criteria has not been determined. Certainly, some individual modules did not fully satisfy instructors who field tested them. The effect of this on the value on the training sequence will be discussed later and tied to the data that was collected.

Training Model Outcomes

Instructor Module Evaluations, Fall, 1975

Five instructors at five universities field tested nine modules during the fall. They wrote about their experiences with each module on the questionnaire, *Instructor Evaluation,* (in Appendix B). Summary statements of their comments follow. On page 64 , Table 1, "Instructor Module Evaluation Comments Converted to Rating Format," presents these data in another format. Finally, conclusions are drawn.

Summary Statements.

1. The overall evaluation across modules centered around three areas:
    a. Readings and materials were essentially good.
    b. There was too much reading and too little activity.
    c. Testing was poor.
2. Most instructors were unable to allow trainees complete freedom in completing modules. One instructor imposed four-week time limits.
3. Most instructors appear to have conferred with students two to three times per module. One instructor indicated 45 minutes per week, while another indicated there was no time for student conferences.

4. Instructions in the instructor's manual were generally seen as adequate. It was suggested that placing the objective on the same page as the enabling activities would be beneficial for both the instructor and the student.
5. Instructor responses indicated that students did require clarification of instructions, etc., primarily due to lack of familiarity with the format of the modules. Instructions were seen as vague and complex for P X. N III was cited as one of the best modules on this aspect.
6. A little over half of the respondents felt the resource materials were adequate. Others felt that they were hard to obtain, were choppy, or did not fulfill the requirements of the module.
7. Responses to post-tests and answer keys were very negative. Instructors felt that questions were vague and required rote memory. The post-tests were seen as punishing to students. One instructor disagreed with duplication of pre-test items on the post-tests.
8. Most responses indicated that the level of presentation was appropriate for most students. Some instructors felt that materials were too advanced and that additional development of content was necessary for adequate understanding of materials.
9. Most instructors found that they were able to keep adequate notes on instructor-student conferences, although responses indicated

that few conferences of this nature actually took place, and those that did take place were of an informal nature, i.e., before or after class periods.

10. All but one evaluator found the resource materials to be easily obtainable.

11. Responses to sequencing of content were, for the most part, positive. P II was criticized for inappropriate sequencing of the positive and negative opinions on career education.

12. Responses to the satisfactoriness of the pre and post tests were primarily quite negative. While one instructor favored objective type questions to "common sense" essays, most felt that the objective questions used were picky, trivial and required rote memorization, or merely common sense. They were seen as too vague to be of assistance in placement.

13. HPTP-produced articles and AV materials were well received for the most part. One instructor stated that students were initially antagonistic to materials but felt it was due to time limits imposed for completion of modules. A couple of materials from P X were pointed out to be dull or outdated.

14. Comments on format were fairly positive. Instructors liked the step-by-step sequencing and cartoons. There was felt to be some problem in the area of choppiness and having to thumb back and forth within the module to tie things together.

15. Instructors who felt that the modules did not meet their instructional needs reported varied reasons for it, such as lack of organization of activities to help structure students' thinking, inadequate development of materials, impracticality of materials or overemphasis on detail.

16. Approximately half of the students consulted with the instructor as suggested by the module instructions. For those students who did not, the instructor felt that it was primarily due to the limited amount of time allotted for completion of the module.

Overview. The overview of these data begins with Table 1, and then to the summary comments; the data are put in the context of actual instructors' comments. This approach will allow you, the reader, to keep in mind the overall nature of the data and also allow particular bits of information to be placed in the context of other data.

The column of modes at the far fight of Table 1 indicates that the modal ratings for two modules were "poor," two "fair," and two "excellent" in terms of conversion ratings. The row of modes at the bottom of the table indicates that four items numbered 2, 3, 7 and 12 were modally "poor" across modules. In like manner, eight items (1, 4, 5, 8, 11, 13, 14 and 15) were "fair" while four items (6, 9, 10 and 16) were "excellent."

At issue is the value of these modules as a whole based on instructor estimates in fall of 1975. It would seem that, from modal values

Module	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mode
P II	2	1	3	2	2	1	2	2	3	2	2	2	2	2	2	3	2
N III	2	3	2	2	3	3	1	3	3	3	2	1	2	2	2	3	3
N II	3	3		2		3	1	3	3	3	2	1	3	2	2	3	3
P X	2			2	1		1	2	3	2	2	1	3	2	1	3	2
P XI	3	1	1	2	2	2	1	2	2	3	1	1	2	3	1	1	1
P VI	1	1	1	3	2	1	1	2	2	3	3	1	2	3	1	1	2
Mode	2	1	1	2	2	3	1	2	3	3	2	1	2	2	2	3	2

TABLE 1. INSTRUCTOR MODULE EVALUATIONS CONVERTED TO RATINGS

1 = Poor
2 = Fair
3 = Excellent

in the rightmost column, only two modules could bear recommendation. From the data in the bottommost row, it would seem that generally the modules were "fair." These data require further analysis. Bear in mind that these latter ratings summarize questions which are about different things and reflect evaluations by different instructors for different modules with different populations and, in some cases, the same instructors at different points in time. Nonetheless, the modules are to be considered as a group.

Converted ratings. For example, the question labeled number 1 in these data is the last one instructors answered. It is, "What is your overall evaluation of this module?" These ratings indicate that three modules were considered "fair" and two "excellent." Overall, instructors apparently thought that the modules were at least "fair." Ignoring P VI, which was rated poor, what were the exceptions to these global assessments? In three cases, instructors said they were unable to "allow trainees essentially complete freedom to complete the module..." (question labeled 2.) In four cases the instructors found the post-test answer key unsatisfactory (question labeled 7). Finally, in four cases the pre-test key was unsatisfactory (question labeled 12). The positive ratings were for questions labeled 9 and 10. Question 9 was about the module task of keeping notes on conferences with trainees.

The actual responses indicate that instructors kept few records of instructor-trainee conferences. Apparently they used an inverted logic; they responded "yes," that it was possible to keep non-existent "adequate" notes (thus leading to the positive ratings). Question 10 referred to the ease of obtaining resource materials. In previous years, materials were difficult to obtain by field test instructors (see Second Formative Evaluation Report, page 62). Obviously, HPTP's efforts to make these materials more accessible paid off in the instructors' estimations.

The exceptions, which modify the global summary statements and the low modal ratings, show some negative outcomes regarding the modules. These were not changed from previous years' outcomes. (See p. 70, Second Formative Evaluation Report.) These exceptions also show that module procedures were not followed in one case (conferences) and were improved in another (material accessibility).

To temper these negative outcomes, two general points should be made here. Question 2 was about freedom to complete the module, and question 9 was about recording instructor-trainee conferences. Use of the modules in some settings make these two steps unattractive due to the time committment required. The Second Formative Evaluation Report, p. 64, indicated that time spent in conferences was small, as well as most of the time spent in instructor-trainee contact. That is, formal classes were not being held at that time. Responses for this year's field tests, indicate that formal classes predominated. Thus "freedom" and "conferences" would have been added burdens. In short, these outcomes are due to different usage of the modules. This resulted in justifiably negative outcomes on these two questions. Such modification in usage might be expected with even less control by the Project.

The second general tempering comment is this: Tests should not be expected to be excellent without field testing. Two things have led to false expectations about the tests in the HPTP modules. The first thing was the naive expectation that item writers could write acceptable items if they tried hard. Several qualifications have to be amended to this statement to make it truer. To begin with, the item writers should have a broad knowledge of the field and of testees' understanding of the field (Wesman, A.G., "Writing the Test Item" in Thorndike, R.L. (Ed.) *Educational Measurement*, Washington, D.C., American Council on Education, 1971, p. 82) Typically, item writers for HPTP have not had a broad background in special education. Doctoral level students, those most obviously qualified in this respect, have had other Project related duties preventing their serving as item writers. In addition, even the best qualified item writers' products must be subjected to field testing to raise their quality. Finally, it is not yet entirely clear in the area of criterion or domain-referenced testing what are supportable ways to analyze the data from field tests. Thus, the production of items has brought a lot of criticism, because it was overly optimistic as conceived and received.

The second thing regarding field testing and writing items is the philosophical zeal with which restricted response (objective) and unrestricted response (essay, problem) items are supported. Evaluatively, restricted response items are easier to analyze. Instructionally, such items often appear trivial. Instructionally, open response items appear desirable, but, for evaluation, they are hard to analyze. Since item writers did not generally have a command of the secondary special education field, they tended to write items which also were restricted in what they tested. Further, early HPTP policy supported restricted response items for analytical evaluative reasons. Field testers obviously do have a command of the secondary special education field. Thus, when confronted with unproven restricted response items and when expecting proven and unrestricted styles of items, they had reason to judge the items to be poor. By their criteria, they are poor items; however, without field test data, no improvement in the items could be made at all.

In sum, the negative outcomes from the *Instructor Evaluations* must be tempered (1) by a consideration of the usage made of the modules and (2) by keeping in mind the problems in writing items as well as the issues surrounding restricted and unrestricted response items. The most serious problems uncovered in this analysis are miscommunications about the functions of field testing the modules and issues of testing which are not unique to HPTP. The question is whether these problems could have been avoided. Probably the issue of miscommunication should never have arisen. Monitoring and constant contact would have helped to combat miscommunication. It seems unlikely that the test items could have been improved given the status of opinions about writing items and given the personnel available. In short, considerable project staff revision would have been required to eliminate these negative aspects found in the *Instructors' Evaluations*.

Summary comments. A quick review of the summary comments reveals that question labeled 15, "Did the module meet your instructional needs?" was responded to fairly negatively. Review of the actual comments shows that instructors were generally prepared "to do a lab." This comment was made regarding module P X, Content Related Resources. Module P II, Career Education, the question labeled 15 brought: "Yes/No. They offered background material. Not practical enough for on-the-job experiences" and "Partially. Complete or thorough set of readings. No real organization or activities to structure student thought." The final comment to module P X is appropriate to most of the responses to all modules for this question, "Module instructions should have forewarned." The expectations for the modules were inappropriate. Module P II, for example, was intended to offer background material. Thus more poor communication resulted in poor reception and misdirected expectations.

Questions labeled 3, 9, and 16 relate to conferences. It is apparent from the summary statements and original comments that student initiated conferences occurred more frequently than instructor initiated ones. These latter were informal and not recorded, as noted earlier. Conferences were seen, within the project, as sources of data for the instructor in supporting and guiding his/her students. This lack of emphasis, therefore, must be considered indicative of problems with the instructional model adopted by HPTP.

Conclusions. The value of these modules is uneven according to these field test instructors. Although the instructors found the readings good, they felt there was too little "lab" or exercise kinds of activities. Part of the reason for these conclusions stems from expectations that these modules would provide something they were not designed to do. The other part of the reason is that HPTP did not communicate the nature of these modules. For example, P II, Career Education, was never intended to be more than background and introductory to this subject domain. Other modules later in the series are intended to provide "lab" experience. The tests are subject to various criticisms and argumentation. Some mitigating circumstances can be brought to defend the tests; however, there remains a need to improve the tests.

Judging by these data, these modules are useful. Nevertheless, instructors were expecting modules more similar to those later in the series. Major staff and content additions are warranted for test improvement.

Module Evaluation and Achievement Data

There were eight modules field tested during the 1975-76 year of the Project. The results of field testing were gathered primarily from three sets of instruments. The first two sets were common to all modules. They were the Instructor Module Evaluation, discussed in the previous subsection, and the Module Evaluation (see Appendix B). Unique to each module were pre and post tests, the third set of instruments. In this subsection we will review the data obtained from the Module Evaluation instrument filled out by trainees, combined with the achievement data from the pre and post tests. This review shows positive outcomes on the eight field tested modules.

A recent study by Stayrook and Corno (1976) and earlier work by Kelley (1967) and Weiner and Kukla (1970) indicate that achievement is a function not only of studying materials but also of subjects' beliefs or attributions about the materials. It should not be surprising that curricular materials do not simply cause achievement, yet this is an assumption implicit in considerable work on achievement. It is the theme of this particular analysis that trainees' achievement is dependent on their attributions, or stated opinions, about the modular materials. This theme extends the rationales of the studies referred to previously.

Attribution theory. Attribution theory was developed by Kelley (1967), Weiner and Kukla (1970) Frieze and Weiner (1971), Kukla (1972), Weiner (1972), Gozali, Cleary, Walster, and Gozali (1973), and Harvey, Arkin, Gleason, and Johnston (1974). It is a logical outgrowth of work on locus of control and related studies on personality and achievement. The central idea of locus of control has to do with the place, site, or locus from which individuals see their lives being controlled. This most radically is stated as having an "internal locus of control" or an "external locus of control" (Rotter, 1966; Lefcourt, 1966). If one sees his or her life as being the consequence of internal locus of control, one sees one's self being able to affect the outcomes of life. If one sees one's self as having an external locus of control, one perceives events outside one's self as determining life's outcomes. Attribution theory is a modification of this radical statement of locus of control.

Attribution theory as presented by Kelley (1967) and Weiner and Kukla (1970) instead states that the individual sees outcomes not just as a function of personal views about one's self but as a function of views about one's self in relation to things around one. Thus a student not only may have feelings about how much he or she can control the world as a rule but also feelings about how he or she can do on a specific piece of work at a particular moment. With respect to HPTP's modules, this means that achievement should be a function of statements students make about the modules and about their relation to the modules. Such attributional statements fall into two classes, personal and material attributes. Personal attributes are statements about ability-luck, effort-task difficulty; material attributes are statements about distinctness, consistency and consensuality. Attribution theory interrelates the two classes of statements.

The attributions studied here were supplied to the trainees in the Module Evaluation form (Appendix B). This procedure allowed control by the investigators of the variety of attributional statements, while at the same time, it secured variation in trainee responses by providing scales on which trainees could answer. The control inherent in this method of development should not have prevented trainees from being able to express the foremost of their attributions. The Module Evaluation form was developed over a period of two and a half years, incorporating commentary which reviewers and earlier trainees made, both orally and by notes. Finally, it was revised to fit it for attributional analysis in the fall of 1975. This revision included the removal of redundant items. These were items which were paraphrases of one another and which earlier trainees answered in the same way as the parallel item. It also included the introduction of some items to cover areas which had only one or two items and which attributional theory suggests are important. These included statements relating particularly to distinctness and consistency of the materials and the ability-luck estimates which the materials call forth from the trainees. In short, although the attributions were supplied, they should have been satisfactory means of expression for trainees, and at the same time did provide the necessary kind of data for attributional analysis.

Attributional analysis has been an outgrowth of personality and achievement, but it also incorporates concepts involving information processing. Kelley (1967) asserts that distinctness, consistency, and consensuality are characteristics of information to be learned, here the curriculum content. For information to be distinct, it must stand out like a visual figure does against a background. It must appear the same time after time, that is, be consistent. Finally, it must result in knowledge which peers see as the same; that is, there would be consensus among trainees. Given that the content of curricular materials possesses distinctness, consistency and consensus as determined fr[illegible] attributions, Kelley (1967) says that students, here trainees, will assimilate the information. Attributional analysis, therefore, characterizes the materials trainees studied in terms of the kind of information in them and in terms of the kinds of achievement the investigators might expect from the trainees.

Procedures. The Module Evaluation form is divided into five areas, i.e., Your Personal Response, The Objectives, The Content, The Assessment Procedures, and The Resources. Since neither the earlier nor this latest revision of the Module Evaluation form was scaled, the investigators could not simply use data from the form. There would be no assurance what the data meant, as measures. Maranell (1974) presents a number of scaling methodologies, any one of which might have been used on the Module Evaluation form. Each was impossible to complete for HPTP. Nevertheless, thematic meaning could be attempted on the Module Evaluation form following Robinson's (1974) recommendations. Essentially this method breaks down the five sections noted above into themes or categories underlying the five and so characterizes the instrument.

The conceptual approach taken was to consider, in Ebel's (1967) terms, the Module Evaluation form to be an individual possessing traits. In such a conceptualization, subjects express statements about a stimulus. For the situation here, trainees used the Module

Evaluation form to make statements about the modules. As noted, their statements effectively related the trainees' self-estimates to perceived characteristics of the modules. Assuming these characteristics included distinctness, consistency, and consensuality and the personal attributions, they should appear in an analysis of the data from the Module Evaluation form. Further, if these characteristics were found, the investigators could expect achievement to be an outcome of studying the content in the modules according to the attributional concept framework.

Factor analysis. Principal components factor analysis will accomplish the reduction of multiple pieces of data to themes or categories underlying the data. For the 75 items on the Module Evaluation form, such a factor analysis was done reducing 54 of the items to independent underlying themes. Six items had missing data, and 15 items did not factor. Factor analysis is an elaborate correlational technique which clusters items together and then can represent the clusters spatially. However, the simplest representation to study is the clusters of items produced. In this case, the clusters are the groups of items from the Module Evaluation form reproduced in Appendix C.

The question is whether the items clustered are thematically related to attribution theory. A factor's theme can be conceptualized from a review of the content of the items clustered on it. This can most simply be done according to Rummel (1970) by taking the highest positive and negative loading items as the theme and then by reviewing items which do not load at all on the factor. (The terms 'load' and 'loading' have technical definitions which are summarized practically by taking them literally. That is, it is useful to think, "A factor carries a 'load', which is measured in terms of 'a loading.'") For example, Items 56 and 57 load positively on Factor One, while Item 60 loads negatively. Items 9 and 31 practically load zero on Factor One. The positively and negatively loading items are about testing and objectives (56), information and modular content (57), and testing and certainty about content (60). The zero loading items relate content and work (9), and the readings and their lengthiness (31). From this review, we can label the theme of Factor One "Testing and Content Specificity." Answers to the items loading on "Testing and Content Specificity" bear directly on trainees' attributions about their ability-luck in answering test items and attributions about the distinctness of content in the modules. Items 9 and 31 load on Factor Two so will be treated in the discussion of that factor. In short, Factor One, "Testing and Content Specificity," has the meaning which is necessary for data analysis under attributional theory.

Factor Two's highest loading items are 8, 9, and negatively, 31. Item 8 relates module topic and trainee interest, 9 - content and usefulness for work, and 31 - readings and lengthiness. The label for the theme of this factor is "Interest and Module Content." Item 9, content and usefulness, was one of the items which loaded zero on Factor One. It fits this theme of Interest and Module Content but has little to do with the "Testing and Content Specificity" theme. Item 31, lengthiness of readings, is part of the theme of Interest and Module Content but not of the theme of Testing and Content Specificity. The theme for Factor Two is related to effort, which is specifically mentioned in Item 1, and to the module making sense (Item 16). There is a practical cast to most items loading on this factor, but clearly these relate to the attribution theory elements effort-task difficulty and to content, interest, sense, or consistency. In summary, this factor and Factor One express attribution theory elements. It is therefore justifiable to view trainees' responses as attributions and to expect achievement related outcomes.

However, before looking at the data from that vantage point, comment on the remaining factors will be made. You will remember that only 54 items were factored, and, of course, we have just discussed 26 items which loaded on Factors One and Two. What happened to those 15 items which did not load on any factors? Why leave out the 15 factors remaining? The 15 items which did not load on a factor simply could not be mathematically accounted for in the factor analysis done. Their mathematical relationships to any of the factors existing in the data were either too small or too inconsistent to show up in the reduction of the data, and they did not show on any other factors either. The 15 other factors, while having relatively large loadings of the items on them, do not amount to very much as factors.

One index which is used to determine the usefulness of a factor is called the "variance explained" and is most easily understood in another form called "the proportion of total variance." To understand these, let us assume that the trainees all thought exactly the same way about the modules; then their attributions would all be exactly the same. In contrast, we assume attributions, under attribution theory, are a consequence of several things, namely ability-luck, effort-task difficulty, distinctness, consistency, and consensuality. Further, we can be reasonably sure that there might be some other things which might result in different attributions, such as different amounts of study, instruction, and preparation which trainees might not be aware of and so not include in their attributions. All these influences cause trainees' attributions to be different or vary. Factor Analysis identifies the total variation which exists in a body of data when it parcels the data into factors. Each factor has some of that "variance explained." Then we make a proportion of variance explained by taking the variance explained for a factor and dividing it by the total variance. You will see in Appendix D, page 179, that fifty-four factors have been printed out along with a column of variance explained, proportion of total variance, and cumulative proportion of total variance. You will note, starting down the right hand column, that seventeen factors account for a .673 proportion or 67.3% of the total variance. You will also note that under the proportion (percent) of total variance that, from Factor 13 down, each factor has less than .020 proportion (2%) of the variance. Finally, if the proportion of total variance is plotted on a graph, a "skree" analysis (See Rummel, 1970) results. This makes a visual analogy between a cliff with the rock rubble at its base (skree) and the graph. It is visually clear that after Factor One, none of the subsequent factors explain very large amounts of variance. These three analytic observations led the investigators to represent only 17 factors in Appendix C. The content of the factors along with our theoretical predictions make it worthwhile to consider Factors One and Two only. The other 15 factors represented are explainable but not worthwhile in terms of how much they cover with their themes.

We now turn to the issue of achievement being an outcome of trainees' relationships with the modules. We have shown via factor analysis that the modules evoke attributions like those expected. The attributional structure provided for trainees was used attributionally in responses by trainees who gave self estimates based on estimates about the modules. We will now present data that the modules evoked statistically consistent or reliable attributions supporting the attribution theory element of consensuality. We will also present some evidence that trainees achieved skills as measured by the pre and post tests. Since the attribution and achievement data are related to the modules, we see this as support for the modules being instructionally effective because they are attributionally sound.

Module evaluation reliability. Calculating the reliability of a measurement instrument is one of the steps in scaling, as was mentioned previously. As scaling was beyond the means of HPTP, neither the evaluation instruments nor the tests had adequate preparation to be able to do the intellectual evaluation on them which is customary. Despite this, intraclass correlation coefficients were calculated on the Module Evaluation form using Ebel's (1967) formulations for the reliability of ratings (attributions here). Any estimate of the Module Evaluation form's reliability would be helpful, and since trainees were making attributions about (ratings on) the modules, an index of the reliability of ratings was appropriate.

Ebel's (1967) coefficient of reliability is called an intraclass correlation coefficient. Essentially, it assumes the variability, mentioned previously in the discussion of the factor analysis, in trainees' responses falls into two major classes: the effect of a variable, like one of those among the attribution theory elements, and error. Now error is assumed to be the random effects of all sorts of inconsistent and conflicting causes. Ebel's (1967) coefficient uses this. It is the ratio of true effects, i.e., that due to a variable, to true effects plus error. The quotient or ratio, true effect over true effect plus error, is the intraclass reliability.

Two points [illegible] interest are central to the meaning of any reliability coefficient. The first point is the underlying structure of the reliability coefficient. The second point is the size of the coefficient. We have already noted part of the meaning of the intraclass coefficient. It relates true effects to true effect plus error. For the intraclass coefficient, there is another condition which adds meaning to this. That condition is that the measuring instrument has a homogeneous theme. In short, by using Ebel's (1967) intraclass coefficient, we assumed that the Module Evaluation provided attributions or ratings having one theme. We already know that was not the case from the above factor analysis. Nevertheless, until the factor analysis was completed, we did not know how the assumed themes would appear in the instrument. (This was one example of the problems which prevented adequate intellectual evaluation of the HPTP instruments.) At any rate, the meaning of an intraclass reliabilty coefficient is the way it relates error to true variability when the data is the result of a single theme.

The meaning of the size of a reliability coefficient resolves to a problem of the meaning of the size of a correlation. Cohen (1969) gives one scaling of size in his discussion of small, medium, and large Pearson product moment correlation coefficients. He gives .10, .30, and .50 as small, medium, and large correlation coefficients and supports these judgements with a lengthy discussion. The point to be made here is the magnitude of the small, medium, and large coefficients is not so large as one might expect before reading Cohen (1969).

Another aspect of the meaning of the size of a correlation coefficient is the confidence we have in the coefficient. Ebel (1967) presents methodology to "estimate the precision" of his intraclass reliability of ratings for incomplete sets. His discussion does not give every detail and requires step by step study. Nevertheless, it agrees exactly with Hays' (1963) "interval estimation procedures" for an intraclass correlation if one assumes Hays' (1963) n equals Ebel's (1967) average k. Both of these methods assume that an intraclass reliability coefficient, as calculated from empirical data, could be inaccurate because of the error attached to the true variation, discussed above. Thus these procedures provide methods by which one can bet on the accuracy of the obtained intraclass correlation. By using these methods, one can find out, regardless of the size of the intraclass correlation, his percentages for betting on the correlation obtained. In short, not only is the size important but also how confident we are that the size is accurate.

By Ebel's (1967) methodology, the investigators obtained an r_I, the intraclass correlation coefficient, of .2038. This was calculated with 290 trainees rating modules on the 75 items in the Module Evaluation. Not all trainees rated on all 75 items; therefore, Ebel's (1967) methodology for incomplete sets was adopted. Using both Ebel's (1967) and Hays' (1963) methodology, we calculated a 95% confidence interval for this intraclass correlation coefficient spanning the interval from r_I of .1501 to r_I of .2589. This means that we can bet that 95 times out of 100 our obtained r_I of .2038 will be between .1501 and .2589. These data indicate that our intraclass reliability coefficient for the ratings using the Module Evaluation form is a small to medium correlation and that we are sure of its size.

The intellectual criteria for reliability require a final analysis of these data, however. Cohen (1969) notes that correlations in the .50 to .60 range are to be found among reliability coefficients. How can this small to medium correlation r_I = .2038, in which we have confidence, mean much as a reliability coefficient? The significance of this reliability coefficient lies in the facts that (a) the ratings using the Module Evaluation had several themes underlying them, yet this size of intraclass correlations resulted; and (b) this coefficient was derived from ratings of several modules; still we are confident of the magnitude of the coefficient. In short, although conditions were bad (Stanley, 1971) for calculating a single form intraclass, or any other, correlation on the trainees' attributions to check their reliability, we have a robust intraclass reliabilty of .2038.

This datum, r_I = .2038, indicates consensus in the attributions made by trainees. Given this and the data that the trainees' attributions fit attribution theory, we can consequently expect trainees to learn in accordance with their attributions about the modules.

Combined point biserial correlations. Unfortunately, it is at this point where the small amount, weak character and complexity of the achievement data begin to obstruct analysis. First we will take some promising-looking data. ModuleEx (for example) was studied by 26 trainees, for which there is achievement data. There were four essay items on the pre test and five essay items on the post test. Each item was scored to indicate whether a trainee had achieved competency or skill in handling the subject of the essay item. Under a competency or performance based training system, on which HPTP modeled its modules, scores indicating that no one achieved competency on the pretest and that everyone did on the post test would be extremely gratifying. In fact, for ModuleEx, this is virutally the case (See Table 2). Twenty-four trainees achieved competency on only the five post test items. One more achieved competency on one pretest item and all the post test items, and one last trainee achieved competency on only four of the five post test items. Furthermore, the attributions on ModuleEx were positive. Those Module Evaluation items which loaded on Factor One averaged 3.46 on a scale from one (low) to five (high). Those loading on Factor Two averaged 3.65. For averages, these are markedly above the converted rating of 3 for "neither" (See Appendix B, the Module Evaluation form) when tested by a t test.

TEST ITEMS

		Pretest				Post test					TOTAL SCORES
		1	2	3	4	1'	2'	3'	4'	5	
Trainees	1	0	0	0	0	1	1	1	1	1	5
	2	0	0	0	0	1	1	1	1	1	5
	3	0	0	0	0	1	1	1	1	1	5
	4	0	0	0	0	1	1	1	1	1	5
	5	0	0	0	0	1	1	1	1	1	5
	6	0	0	0	0	1	1	1	1	1	5
	7	0	0	0	0	1	1	1	1	1	5
	8	0	0	0	0	1	1	1	1	1	5
	9	0	0	0	0	1	1	1	1	1	5
	10	0	0	0	0	1	1	1	1	1	5
	11	0	0	0	0	1	1	1	1	1	5
	12	0	0	0	0	1	1	1	1	1	5
	13	0	0	0	0	1	1	1	1	1	5
	14	0	0	0	0	1	1	1	1	1	5
	15	0	0	0	0	1	1	1	1	1	5
	16	0	0	0	0	1	1	1	1	1	5
	17	0	0	0	0	1	1	1	1	1	5
	18	0	0	0	0	1	1	1	1	1	5
	19	0	0	0	0	1	1	1	1	1	5
	20	0	0	0	0	1	1	1	1	1	5
	21	0	0	0	1*	1	1	1	1	1	6
	22	0	0	0	0	1	1	1	0*	1	4
	23	0	0	0	0	1	1	1	1	1	5
	24	0	0	0	0	1	1	1	1	1	5
	25	0	0	0	0	1	1	1	1	1	5
	26	0	0	0	0	1	1	1	1	1	5
Combined point biserial correlations		.00	.00	.00	.71	.00	.00	.00	.71	.00	

$$r_p = \frac{\overline{Y}_1 - \overline{Y}_0}{S_y} \sqrt{\frac{N_1 N_0}{N^2}}$$

See Bruning and Kintz (1968) for the computation of r_p, the point biserial correlation.

See Lord and Novick (1974) for its theoretical development.

Most authors seem to use N^2 rather than the $N(N-1)$ which Bruning and Kintz (1968) use.

The discussion represents $N_1 N_0 / N^2$ as S_{RW} and S_y as S_T; $\overline{Y}_1$ as M_R and $\overline{Y}_0$ as M_W.

} the only variation in the total scores

*the only variation in pre or post test item scores

TABLE 2. MODULEEX COMBINED POINT BISERIAL EXAMPLE

In short, ModuleEx has all the appearances of an ideal module by competency or performance based standards as well as by the attributional theory standards adopted.

ModuleEx, however, is an example of the complex problems attending data analyzers in the area of DRT (domain referenced testing) or CRT (criterion referenced testing), the form of testing attempted by HPTP. It should be noted that DRT or CRT is not intrinsically required in competency or performance based instruction. Nevertheless, test data from norm referenced tests, the most common form of tests, are usually interpreted like DRT or CRT tests when used with this form of instruction. For this form of instruction and the DRT or CRT form of testing, the results from ModuleEx are, to repeat, gratifying. They clearly show the effects of instruction and positive attributions by trainees. They also defy analytic procedures.

Haladyna and Roid (1976) discuss several indices of item quality. One, in particular, was selected for use with the data discussed in this section. Another, the pre to post difference index or Cox-Vargas index, used to analyze data in Appendix A, requires identical items on pre and post tests. For most of the modules, this situation did not exist. Another index presented by Haladyna and Roid (1976) does not have this limitation. It is the combined point biserial correlation. This index is calculated for each item with pre and post tests combined. In essence, it correlates the number of right answers to an item with the total scores that subjects had over all pre and post test items.

The point biserial correlation, r_p, is the correlation between something scored in two categories - right-wrong, male-female - and something which has continuous scores (e.g. a total). The r_p can be understood from its calculation formula, $((M_R-M_W)/S_T)S_{RW}$. M stands for mean or arithmetic average; S is for standard deviation, a mathematical representation of variation in measurements. R, W and T are labels meaning right, wrong and total. One major use of the point biserial is to find the mathematical relation between testees' answers to an item on a test and testees' total test scores. To get M_R, one finds every testee who got an item, say 7, correct. Then one takes each of these "correct" testees' total scores on the test and gets an average of these totals to make M_R. To get M_W, the mean of total scores for those who got item 7 wrong is found. M_W is subtracted from M_R and compared to (divided by) S_T, the standard deviation of total scores. The final step is to multiply by the standard deviation of the right-wrong answers, S_{RW}, which makes r_P have values between -1 and +1. Thus r_P is $[((M_R-M_W)/S_T)S_{RW}]$.

This index, r_P, becomes the combined point biserial correlation when the pre and post tests for trainees are combined. The combined total of pre and post test scores for trainees increases the variation in total scores and hence the standard deviations S_{RW} and S_Y. You will note in Table 2 that all 26 trainees failed to achieve competency on item one (all zeros in the column for item 1). There is no variation, so S_{RW} is zero for item 1. When all the items for a test have no variation, neither does the test. Thus, if the scores are all zeros, which is almost the case with the ModuleEx pretest, or all ones, which almost happened with the ModuleEx post test, then there is no variation in total scores for pre or post tests. By combining pre and post tests, score variation is increased. Usually the pre and post tests are not so completely "none" or "all," so combining tests fattens up the variation for the point biserial and makes it possible for the r_P to be closer to -1 or +1. When the combined r_P is close to +1, it indicates that items correlate with the total combined scores.

Conceptually, you can see that if pretest items do not correlate with the combined total scores and that post test items do correlate, then instruction probably made the difference. However, with no variance, the effect of instruction which is clear in Table 2 for ModuleEx test items cannot be shown by using the combined point biserial correlation because if S_{RW} is zero, so is the correlation! This case, ModuleEx, obviously links instruction and attributions; yet the mathematics which are so essential in cases which are not so obvious, fail here to support or even to modify these obvious links.

So much for complexity. Let us take a moment to examine the character of ModuleEx items. As noted, these were essay items. The major problem with such items is getting between graders or within grader consistency. Two graders seldom agree, and one grader will seldom grade items the same way twice. In this latter case, the situation-knowledge of previous achievement, personalities, effort, ability and so forth affects the grader's marks. It is not unreasonable to expect that in the competency based format adopted, instructors and teaching assistants would have situational pressures, particularly pre and post test instruction ones, to influence their grading. This is not to imply conscious bias. but it is to question the ability of graders to be able to overcome unconscious factors. The model answers supplied for these ModuleEx items were specific as to the elements to be listed and discussed. Nevertheless, the grades, i.e. competent-not competent, do not specifically reflect the elements on which they are based. Thus the character of ModuleEx items is weak with reference to objective criteria.

Conversely, items on other modules were objective but were weak because of problems such as those discussed in the section on the Instructor Evaluation form. Specifically, judgements of triviality and poor item writing suggest weaknesses among the objective items.

Finally, let us consider the amount of data. P II is an example. There were only 26 trainees who studied with this module. Further, these were the first to use the module. Test development and analysis requires considerably more subjects to supply data, and it requires the recycling of the tests for revision and retesting. These additions are necessary from several viewpoints of intellectual evaluation. Some have been discussed; other more complex ones become apparent with study of articles on test development. To be blunt, statistical analysis of tests requires appropriate and large quantities of data. The modules had too little data.

Linking attribution data to achievement data for HPTP data is seriously impeded by the complexity, character and amount of achievement data. We have reviewed some of the most prominent problems in working with the HPTP data.

Achievement. The combined point biserial correlation is an index of item-criterion correlation. That is, it indicates the relationship and contribution of an item to the test score. When that criterion or test score is the combined pre-post test totals, as was the case here, it gives an index of the worth of items for differentiating between post test trainees who should know modular material and pretest ones who do not. In short, the point biserial correlation is useful to test the quality of items, and it is useful to indicate achievement of competency when used in the combined form (Haladyna and Roid, 1976; Henrysson, 1971). The point biserial should work well with combined pre and post test scores because we expect the mean item difficulty to approximate .5000, a condition for the point biserial (Henrysson, 1971). Thus, there are three reasons for using the combined point biserial correlation: (a) it is an item index, (b) it can be a test index, and (c) the combination of pre and post test scores should provide appropriate conditions for using the index for CRT (criterion referenced testing).

Return to the example, ModuleEx. In that case, the mean item difficulty is .56. (Items 1-3 had zero difficulty and item 4 had a difficulty of .04 (1/26). Items 1', 2', 3', and 5 had difficulties of 1.00, and item 4' had one of .96 (25/26)). This is what we might expect under a competency based system using CRT. We can even expect pretest item difficulties to be higher and post test ones lower and still expect the mean difficulty to be around .50. This would be the ideal case for using the combined point biserial correlation, for, as we have seen with ModuleEx, the index is zero when there is no variation.

Certainly ModuleEx should be the rarity, with test items which are unproven.

To account for problems in variation like those with ModuleEx, it becomes necessary to examine the item difficulty along with the combined point biserial correlations. "Item difficulty" is the number of people who got an item right divided by the number who were tested with that item. If 5 people were tested with an item, and 4 got it right, 4/5 or .80 would be the item difficulty. Note, even though this is called "item difficulty," it gets larger as more people get the item right. The item difficulty is what S_{RW}, the standard deviation of the right-wrong answers, is based on. Thus, the combined point biserial correlation is dependent on item difficulty. This is what makes it unnecessary in most cases to refer to the item difficulty. However, where item difficulties are extreme, one must review them as well, since S_{RW} becomes small, making the correlation small even where conceptually it is high.

Table 3 presents mean combined point biserial correlations and mean item difficulties derived from nine field tests of HPTP materials. For all of the materials, except N III-Kentucky and P II, there is an increase in the combined point biserial correlation of items from pre to post test. Assuming we were confident that the data had no extremes of variation or other difficulty, we could be satisfied that

Module	Mean Combined Point-Biserial Correlations		Mean Item Difficulty		
	Pretest	Post Test	Pretest	Post test	Combined
N III KU	.2651	.3677	.6216	.6864	.6466
N III KY	.2192	-.0119	.5513	.9756	.7154
N IV fall	-.0370	.4267	.6362	.7516	.7073
N IV spring	-.0855	.6232	.7602	.2557	.4238
N V	-.1170	.4956	.5095	.6714	.6020
P II	.0000	.0000	.0096	.9923	.5556
P VI	.1621	.4483	.6326	.6519	.6431
P X	.2293	.4079	.6812	.7484	.7313
P XI	.1136	.2970	.7511	.7115	.7382

TABLE 3. TWO PAIRS OF INDICES FOR PRELIMINARY ASSESSMENT OF ACHIEVEMENT

achievement resulted from trainees using the materials. We know differently; hence, we examine the mean item difficulties.

For the most part, the item difficulties indicate that the value of the difficulties increased on the post tests. Thus support for achievement is found here. Nevertheless, there are inconsistencies with the biserial indices. N IV-Spring difficulty indices do not agree in directions with the point biserial indices; P XI difficulty indices do not agree with the point biserial indices.

From looking at the biserial correlations and the difficulty indices for the data from nine field tests, we have found four sets which are unclear in their implications: N III-Ky., P II, N IV-Spring, and P XI. N III-Ky. data had so little variation in the post test scores (most were correct) that the post test point biserial went to zero. Thus, the direction of the difficulty indices (.5513 to .9756) is indicative of achievement.

P II data is very similar to ModuleEx data. Everyone got items wrong on the pre test and everyone got items right on the post test. The lack of variation made the point biserials zero, but achievement appears in the item difficulties, which go from .0096 to .9923. Note that the combined item difficulty for P II was .5556, the goal in this form of testing.

N IV-Spring data appears incomprehensible. As the difficulty indices show, post test difficulty was low, indicating few trainees showed competency on those items. One might guess that the reason would be that the post test items were on subject matter trainees had not learned. In fact, two things happened. There was disappointment with N IV tests, and this led to non-test measurement for assessment of competency. Thus the N IV-Spring data mean nothing in terms of competency.

Finally, P XI data have moderately high difficulty indices in the .7000's. This would tend to reduce variation. The point biserials indicate an increment in correlation with combined scores, but there is a decrement in difficulty indices. The point biserials are small due to the difficulty indices we know. Could they, nevertheless, indicate in fact what their relative sizes indicate, acquisition of competency? Further analysis would be required to answer this question. The present conclusion must be one to question rather than decide on these outcomes.

In sum, these data, presented in Table 3, present foundation for preliminary belief that the field test use of HPTP materials led to noticeable acquisition of competency. This conclusion is based on complex, weak and small amounts of data.

Conclusion. This application of attribution theory to the analysis of competency based, criterion referenced testing to evaluate the achievement of competencies has been frustrated by problems in the test data used to indicate competency. Nevertheless, a picture remarkably consistent with attributional theory emerges in review of the data. Essentially, this picure shows trainees learning from materials to which they attribute informativeness. Despite some criticism of HPTP's modular materials at field test sites, trainees apparently found value in the materials. This conclusion is supported by positive attributions from items loading on the two factors providing major themes for the Module Evaluation form. In consequence and in spite of complex, weak small amounts of data, seven of nine field tests of HPTP materials show preliminary indications of achievement of competencies resulting from trainees' work with the materials.

Further work on these and new data on the modules is required. Adherence to attribution theory constructs which recognize trainees' part in learning is recommended.

The Evaluation Model

In the section on sources of goals for HPTP, early in this report, the goals noted pertained primarily to the products to be expected from the Project. Little was written down about goals for the Project itself. A number of points will be made here about Project and product goals.

Project Goals

First, it is evident that Project goals were primarily product oriented. Administrative decision making, staff composition, and staff orientation were geared to produce the curriculum contained in the modules.

Secondly, but less evidently, it was a Project goal to support a number of doctoral and masters level students in their work at the University of Kansas. Such support was not only instructive but, just as importantly, provided them with experiences valuable for their later work.

Thirdly, a case could be made for HPTP to support training at other universities. In effect, the field test procedures did this, while at the same time contributing to the data base on the modules. This is an unexpressed goal for HPTF so that the consequences are justifiably called side effects.

Fourthly, summative evaluation, along with formative evaluation, was implicitly a goal for HPTP. Nevertheless, preparation for summative evaluation would have required at least another year. This is partly due to pressures to produce and formatively evaluate modules and partly due to technical issues surrounding competency based and criterion referenced testing.

Fifthly, it may have been a Project goal to test the Competency/ Performance Based Teacher Education (C/PBTE) paradigm. While there is some latitude in interpreting what "test" means, no significant work was done on what would have had to be research of this kind.

Actions and outcomes relevant to the first of these five goals of evaluation have been detailed earlier. The lessons in management and hiring practices qualify as successes for members of the staff. Benefits to the Project resulted. Nevertheless, two areas suffered neglect. First, turn around for modules ready for revision after expert appraisal and field testing was slow and not clearly predicated on rules for response to the criticism. Secondly, dissemination criteria tended to overshadow vestiges of rules which did exist for revision. This second area will be discussed later.

The most potent example of the slowness in revision after criticism is the revision process with Modules N III, IV, and V (*World of Work, Home and Self* and *Communtiy and Society*). Three months during 1975-76 were spent in meetings and derivative work to settle on a revised format and modified content for these three related modules. These were some of the earliest modules completed, so they had been appraised and field tested. Data presented in the Second Formative Evaluation Report indicate that these modules did not present "new material" and that there were mismatches among objectives, activities, and tests. In addition, the staff had noted the trainers' and trainees' desires for experiential activities and had found problems of redundancy between these and other modules. As noted, this work took three months to complete for these modules alone. Revisions in other modules were less drastic and took less time.

HPTP helped support a number of doctoral and masters level secondary special education trainees during its existence. Below is a table detailing that support and the outcomes.

	Approx. Support	No. Supported	No. Graduated	Programs in Progress
Doctoral	$99,942	6	3	3
Masters		11	4	5

TABLE 4. SPECIAL EDUCATION STUDENT TEACHERS WHO RECEIVED FUNDS FROM HPTP

Of the three doctoral students who remain to graduate, one was first a masters student and will in all likelihood graduate in another year and a half. The other two are in the midst of dissertation work and should graduate this year. Of the eleven masters students, two will probably not graduate. One of these people was a courageous handicapped individual who became overextended by attempting a masters. The remaining five will probably complete their final work. All these individuals contributed to the Project with their work on content, production, or evaluation.

The support which HPTP provided for training at the universities and colleges is undeniable, yet hard to clearly specify. Courses were already designed and being taught where the modules were tested. The selection criteria for the field test sites required this. Nevertheless, the modules which were tested must have provided, as side effects, some new material or supplanted other material. Whichever was the case, HPTP materials were studied and discussed by approximately 300 students at six universities or colleges. Only achievement and rating data, detailed earlier, were obtained from the secondary special educators. Other side effects may have resulted or may result, but data could not be collected. Undoubtedly, the supportive function of HPTP's field testing resulted in additional unidentified side effects.

Formative evaluation was a goal and was used for three years of the Project. This was based on the design illustrated in Figures 4-6. The design served adequately for decision making. It was not so complicated that it could not be followed in practice nor so ineffective as to provide no support for decisions. Nevertheless, it was slow in

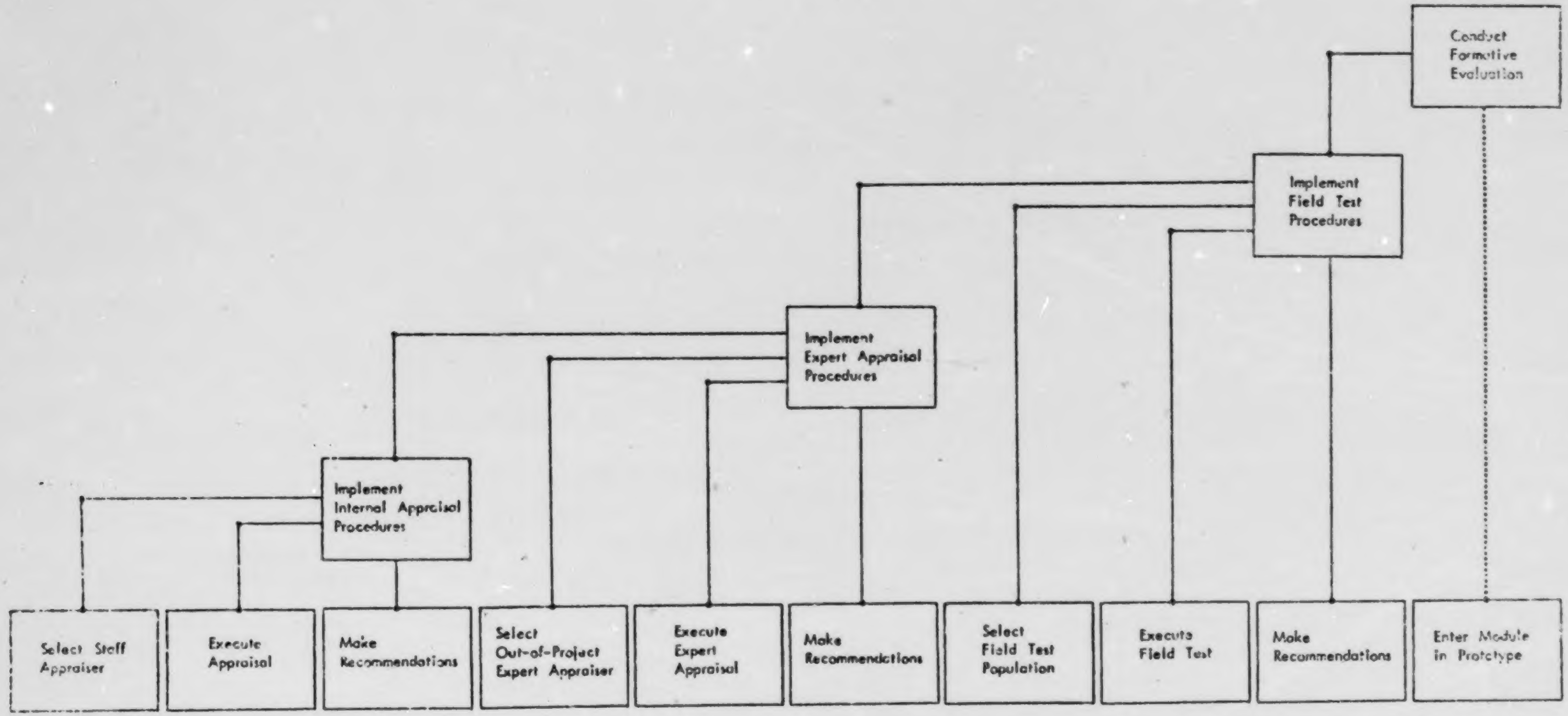

Conduct Formative Evaluation
Implement Field Test Procedures
Implement Expert Appraisal Procedures
Implement Internal Appraisal Procedures
Select Staff Appraiser
Execute Appraisal
Make Recommendations
Select Out-of-Project Expert Appraiser
Execute Expert Appraisal
Make Recommendations
Select Field Test Population
Execute Field Test
Make Recommendations
Enter Module in Prototype

FIGURE 4: THE FORMATIVE EVALUATION MODEL

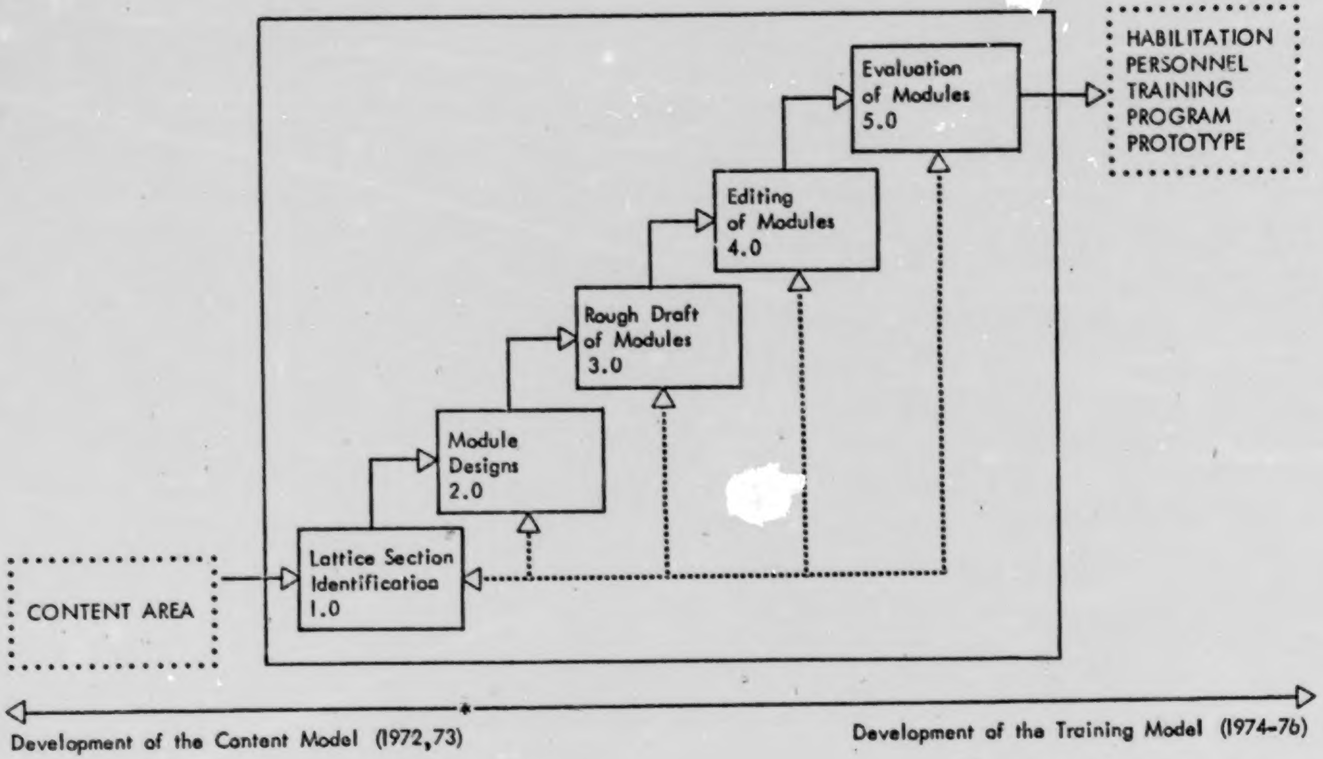

CONTENT AREA
Lattice Section Identification 1.0
Module Designs 2.0
Rough Draft of Modules 3.0
Editing of Modules 4.0
Evaluation of Modules 5.0
HABILITATION PERSONNEL TRAINING PROGRAM PROTOTYPE
Development of the Content Model (1972,73)
Development of the Training Model (1974-76)

FIGURE 5: THE MODULE PRODUCTION PROCESS INCLUDING FORMATIVE EVALUATION STEP 5.0

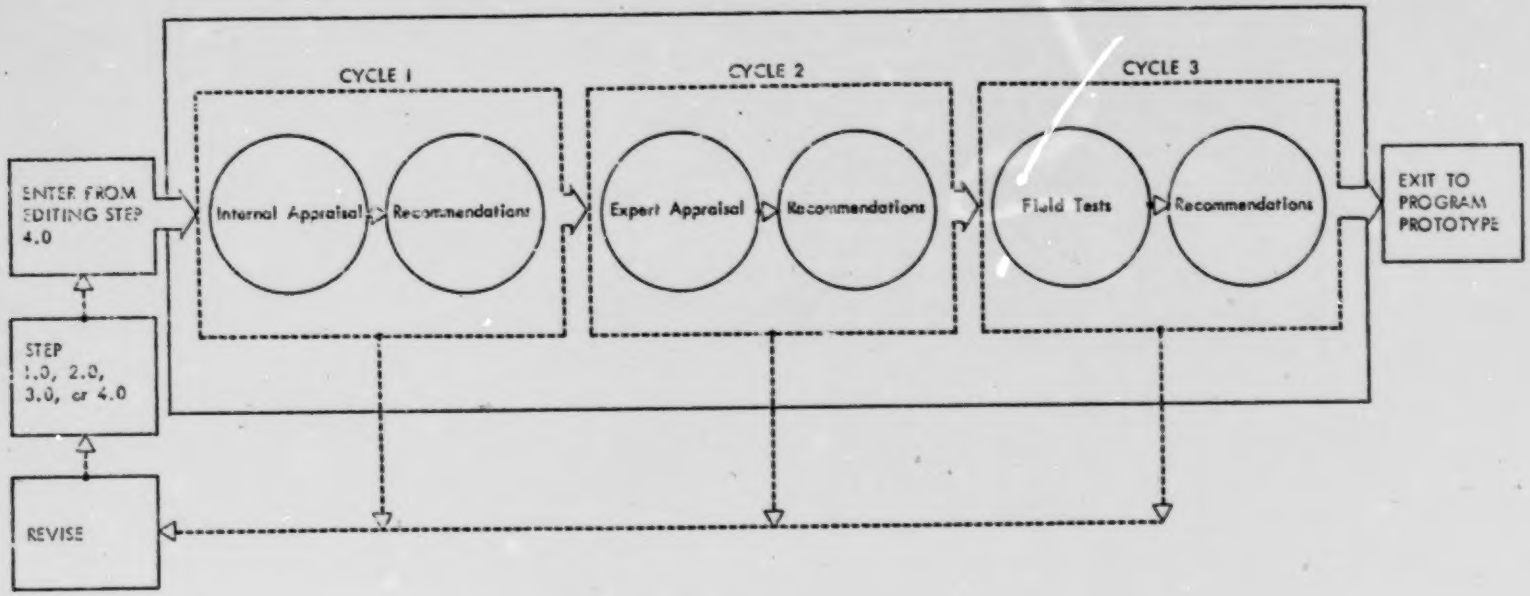

CYCLE 1
CYCLE 2
CYCLE 3
ENTER FROM EDITING STEP 4.0
Internal Appraisal
Recommendations
Expert Appraisal
Recommendations
Field Tests
Recommendations
EXIT TO PROGRAM PROTOTYPE
STEP 1.0, 2.0, 3.0, or 4.0
REVISE

FIGURE 6: THE EVALUATION-REVISION CYCLES IN STEP 5.0

producing results. This was partly because the procedure for obtaining the data was slow, partly because the procedures for reducing the data were interrupted by personnel changes, and partly because application of the data to revision was delayed by production of new modules.

Summative evaluation is a companion of formative. As used here, it should cover intellectual and social criteria. Among intellectual criteria should be reliability and validity of measuring instruments. At present, there are a number of controversies surrounding how to arrive at the reliability of a criterion referenced test and how to place testees based on the "reliable" outcome of such data. These are not issues about competing methods or themes but about what methods there are. Lord and Novick (1968), p. 63, point out that "...the reliability of a test is just its validity with respect to a parallel test." Hively, Maxwell, Rabehl, Sension, and Lundin (1973) offer some methodology for developing items in the same domain, i.e. parallel tests. Durnin and Scandura (1973) modified this methodology to place a conceptually satisfying method in reach. Whether generally satisfying empirical results can be obtained will take time. The gist of Durnin and Scandura (1973) is that "algorithm [or rule] based technology" can generate psychologically real domains, whereas the Hively, et. al. (1973) procedures provide logically and psychologically incomplete specifications for domains. In short, developing psychologically parallel domain referenced tests is a serious theoretical and practical problem preventing the collection of clearly supportable reliability data.

An exploration into the validity of our criteria was done using what has become known as the Cox-Vargas index. It is detailed in Appendix A: Adjunct Research. As you will see in that section, there are serious problems in using experienced subjects for setting test criteria.

Intellectual criteria were also used in the attribution analysis of responses to the modules reported in the Training Model section. To summarize here, this analytic framework incorporates the assumption that trainees will work on and achieve from the modules in accordance with their personal estimates of how well they will do. If a trainee estimated that he/she had the ability and that the materials could be learned from, he/she would do better than someone who made negative estimates on one or both of these. The results are encouraging. Additional work related to such interactionist and even noncausal models is needed.

The social criteria for summative evaluation include at least usefulness and effectiveness. Instructor and module evaluation data indicate problems with the usefulness of the modules. Nevertheless, instructors for the most part (8 of 9) requested additional use of the modules they had field tested and at least 43 other instructors (as they represented themselves to the Project) requested use of the materials during our dissemination efforts. The data indicate that the modules field tested were instructionally effective. Trainees learned from them based on our testing. This obviously is subject to the technical qualifications mentioned previously. Trainees generally reacted positively toward the modules. The evaluation data give indications of usefulness and effectiveness.

On the other hand, publisher response, another social evaluation criterion, was not so positive. Initial reactions of publishers indicated that the array of field test modules was too expensive to publish for the "thin market" open to the materials. The details are given in the dissemination section. With respect to evaluation, perhaps the best design would have been to build in a collaboration with a congenial publisher. The risk would have been that publishing could have overbalanced other factors such as the content or training approach adapted. The benefit could have been a product which would have been tested in a form more nearly like that which might be published. As it turned out, the modules format did not meet publishing criteria in the thin market.

As detailed previously, summative evaluation was implicit in the adoption of formative thinking for HPTP. The intellectual criteria for summative evaluation have proven to be beyond reach, given the state of the art in criterion-referenced and performance or competency-based testing as used by HPTP. Summative data indicate that field users were anxious to use the materials despite problems. On the other hand, the mass of materials were unpublishable by editorial standards communicated from publishers. Thus summatively, most intellectual criteria simply could not be met and social criteria were satisfied in one case but not another.

A final goal for the Project may have been to test the Competency/Performance Based Teacher Education (C/PBTE) model incorporated in the modules. This goal was unclear among Project members and contributed to weaknesses in evaluative efforts. The fact is that this goal is a research goal. Without a doubt, all the Project resources would have had to have been devoted to this goal in order to meet it. Given the present technical problems mentioned in regard to reliability and validity, it is likely that the outcome of the research would have been unsatisfactory. The Project was not prepared intellectually nor by its staffing to accomplish the goal of testing C/PBTE.

Product Goals

The product goals for HPTP were specified in detail. The manner of formatively evaluating production for recycling was part of this specification and is detailed in Figures 4 through 6. Decision making was relatively well supported for the work on the modules. Two exceptions are noted below.

As noted in the history, considerable difficulties surrounded the processing sequence for modules. In particular, the review procedures became overelaborate and thereby unworkable. There was a retreat from this over specificity to the design presented here. Unfortunately, neither the very specific nor the workable design could be organized to keep the revision rate at a compatible level to the production rate. This was partly due to the fact that the number of reviewers and their commitments prevented their meeting the workload of review. It also partly was due to the complexity of reviewing and then revising "completed" modules. The revision rate was uneven.

High among the product goals was a Competency/Performance Based Teacher Education paradigm. Data from field test instructors and trainees indicate that they felt competency and performance were slighted. Part of this dissatisfaction probably was due to inappropriate expectations for the particular modules field tested. Part of it also is traceable to efforts to provide "higher level" activities (in reference to the Bloom, Engelhart, Furst, Hill, and

Krathwohl, 1956, taxonomy). The effects of these efforts were that staff placed more emphasis on reading and writing to "tie things together" and less on objective criteria for the activities. Neither of these outcomes appears necessary; however, psychological and sociological models of learning and teaching which have had to have been brought in. Further, competencies which are primarily inferential would have had to have been reduced in their importance in favor of more operationally defined competencies. Neither adopting psychology or sociology, nor making competencies more concrete would be simple steps. The options for meeting the C/PBTE goal were available but difficult.

Two exceptions to meeting product goals have been discussed in this section. No consideration of the rule, which these exceptions prove, for meeting product goals was made because of data available elsewhere in this report. Such data indicate that meeting goals was closer to the rule. The product goals discussed here have had high priority but were difficult to meet in satisfactory ways.

Overall, the evaluation model worked. Nevertheless, the results leave things to be desired. Some additional reasons for this state will be considered in the Discussion and Recommendation section of this report.

Summative Evaluation

The summative evaluation model laid out in Figure 7 is the basis for this report. It is presented to complete discussion of the evaluation efforts and to summarize summative rationales for the Project.

This model is centered around the four major tasks set for the Project and around the goals set by, actions initiated in, and outcomes realized through Project work. No comprehensive model, adapted from major evaluation theorists, was used over the whole Project. This decision was based on the facts that evaluation models are not so comprehensive philosophically as to cover the peculiarities of this Project, that a fully systematic summative evaluation was not planned in time to prepare the necessary instruments, and that some issues brought up by such models could be treated within the framework of this model. Evaluation models available were difficult technologically or philosophically and so were either unusable or inappropriate, or were so relatively simple as to be usable, where relevant, with ease. Thus models requiring considerable intellectual outcomes relating to reliability and technical generalizability, for example, were not implemented. Models, such as Scriven's (1974) checklist, were searched for relevant topics. However, the economic philosophy in this checklist was less appropriate for the Project than the idea of a social pact. This pact was made to solve

	PROJECT TASKS REVIEWED IN THE SUMMATIVE REPORT			
REVIEW TOPICS	DEVELOP CONTENT	DEVELOP TRAINING	DO EVALUATION	DISSEMINATE
NEEDS AND OBJECTIVES TO MEET UNDER TASKS	For Trainers, Trainees, Students	For Developers, Trainers, Trainees, Students	For Project	For Trainers, Trainees
PROJECT ACTIONS TAKEN TO COMPLETE TASKS	Developed Lattice	Developed Training Sequence	Field Tested Modules, & Did Attribution & Adjunct Studies	Wrote Letters, Presented At Meetings, Contacted Publishers
OUTCOMES RESULTING FROM ACTIONS	Received Positive Criticism, Had Lattice As Conceptual Foundation	Accumulated Data: Instructor Evaluation, Module Evaluation & Achievement Data	Used Data For Formative Based Decisions, Did Summative Evaluation, Demonstrated Accountability	Received Positive Responses At Presentations & From Correspondence
DISCUSSION OF OUTCOMES	Content Was Justified Intellectually & Socially	Field Tested Materials Effective Under Attribution Approach, Materials Furnished	Slow Recycling, Some Intellectual Criteria Possible To Meet, Social Criteria-Positive Outcomes	Materials Available For Publishing, Demand Good, Dissemination Incomplete
RECOMMENDATIONS	Lattice Needs Validation, Add Feedback To Lattice	Field Test New Format, More Psychology & Sociology Of Training Needed	Better Technology Needed, "Causality" An Inappropriate Assumption	Publisher Contract Earlier, Experiment With Salability Criteria

FIGURE 7: SUMMATIVE EVALUATION MODEL

The Dissemination Model

The dissemination model presented in Figure 8 follows United States Department of Health, Education, and Welfare -- National Institute of Education and Office of Education -- "Copyright Guidelines." It is not unique to this Project. As noted in Figure 8, there are two branches of effort involved: publicizing the materials and, on obtaining the proper protection and clearances, publishing the materials.

Publicizing the modules was a continuous process in the last two years of the Project. It included writing to people about the module sequence, both on the initiative of Project staff and at the request of the interested parties. It also included presentations at several meetings of educators of the handicapped.

Writing at the initiative of Project staff members included contacting Expert Appraisers, Field Test Instructors, plus notifying people on the Project's mailing list, developed in the 1975-76 year, with brochures. Expert Appraisers included during 1975-76:

Dr. Robert Fuchigami
Ms. Patricia P. Kells
Dr. Gary M. Clark
Dr. Joseph Eisenbach
Dr. Charles Kokaska
Dr. Oliver Kolstoe

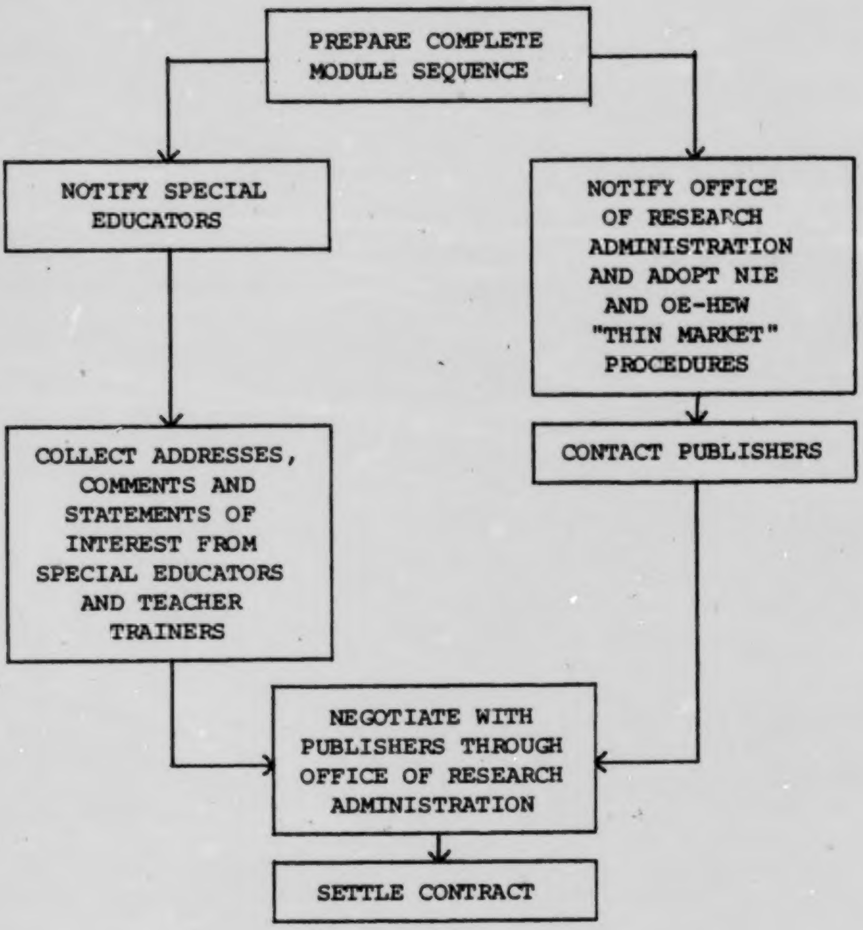

PREPARE COMPLETE MODULE SEQUENCE
NOTIFY SPECIAL EDUCATORS
NOTIFY OFFICE OF RESEARCH ADMINISTRATION AND ADOPT NIE AND OE-HEW "THIN MARKET" PROCEDURES
COLLECT ADDRESSES, COMMENTS AND STATEMENTS OF INTEREST FROM SPECIAL EDUCATORS AND TEACHER TRAINERS
CONTACT PUBLISHERS
NEGOTIATE WITH PUBLISHERS THROUGH OFFICE OF RESEARCH ADMINISTRATION
SETTLE CONTRACT

FIGURE 8: THE DISSEMINATION MODEL

Dr. Ted R. Whiting
Dr. Lotar V. Stahlecker
Dr. Donn Brolin
Dr. David Gardner

Field test instructors, whose support was essential, were:

Dr. Bruno D'Alonzo
Dr. Rex Schmid
Dr. Lotar Stahlecker
Ms. Pat Cappozzoli
Dr. Patricia Cegelka
Dr. Stuart Schwartz
Dr. Gary M. Clark
Dr. Patricia Sitlington
Ms. Diane Wimmer

There were 87 people on the mailing list who were sent brochures describing the modules and the Project. This was done early in the spring of 1976. A number of additional contacts resulted. Dr. Kokaska, for example, wrote indicating that several school districts in southern California would be interested in the modules. Subsequently, some of their districts wrote for details. Dr. Edward Skinski, located in Washington, D.C. consulted for Dr. Gregory Little in Palmer, Massachusetts, and recommended our materials. Dr. Little expressed strong interest in using the modules.

Inquiries from interested individuals were the basis for half the mailing list. These people inquired about every aspect of the Project and included one proposal, which was accepted, from Ohio's HELPS. This proposal was to modify HPTP modules for inclusion in a computer-based information system. Modules so stored were to be accessible on a statewide basis for training. Four modules and introductory material were modified at the University of Kansas by John Kregel, a former masters student, and three other people, all four paid by Ohio's HELPS. James F. Daiker, with Ohio's HELPS, was liason for this externally initiated publicizing of HPTP materials.

Dr. Wayman Shiver at Miles College in Birmingham, Alabama, also inquired at length about adapting and adopting HPTP materials for use with Teacher Corps trainees. Contact with Dr. Shiver included his visiting the Project, with the outcome to be the use of our materials on publication. Dr. Shiver had surveyed several bodies of materials and found the Project's, even at the secondary level, to be modifiable to the needs of his elementary level Teacher Corps program.

Publicizing also included presentations at general meetings of special educators. This included presentations at the State meeting of the Council for Exceptional Children (CEC) held in Kansas and a similar meeting in Oregon during the 1974-75 year. It included a general presentation of the Project at the National Teacher Education

Workshop on Vocational Education for Special Needs Students with a display and discussion of HPTP products in Champaign, Illinois, on the University of Illinois campus. Finally, HPTP participated as an exhibitor at the National Convention of CEC in Dallas (1973), New York (1974), and Chicago (1976).

It was estimated that 200 to 300 persons attended the state CEC meetings, with ten to fifteen actually attending presentations given.

At the Champaign workshop 62 active participants were audience to the general presentation made about HPTP. The emphasis in the workshop was on linking vocational and special education programs within universities across the nation. Feedback on the workshop presentation was generally positive.

Response to HPTP materials at this workshop was especially intense and positive. Twenty persons asked for brochures on the materials after the above presentation. Six people discussed the modules for periods of five to thirty minutes. One person spent thirty minutes discussing the materials displayed and nearly missed the next morning meeting. Another returned later in the day for twenty minutes. The remaining four spent five to ten minutes discussing the materials and giving predominantly positive comments. Exceptions were disagreement with the ordering of topics and, in one case, a request to cover a topic not in the modules on display. These comments were recorded at the time they were given and were, with the exceptions noted, aggressive but positive.

As an exhibitor at the April, 1976, CEC convention in Chicago, HPTP gained considerable exposure for its modules. Of 1,000 brochures, 600 were taken the first exhibit day. The remainder were picked up early the second day. There were 11,221 registrants, and traffic among the 297 exhibits remained high this second day. Traffic tapered off the third day, but several people returned to discuss the modules at length, something they did not want to do in the crush of the previous two days. These people included a psychologist, several special education program coordinators, and teacher educators. The figures above and those to follow come from *Resource*, volume 6, number 2, for June, 1976; it is a CEC newsletter for "marketers of education materials for exceptional children." Attendance breakdown by professions included primarily those listed in Table 5 on the following page.

Teachers	27.6%	3,097
Administrators	16.1	1,806
Students & Student teachers	28.0	3,142
Teacher educators	12.7	1,425
Psychologists	3.1	348
Others	12.5	1,403
	100.0%	11,221

*TABLE 5. PROFESSIONS OF REGISTRANTS AT THE 1976 CEC CONVENTION**

*based on data supplied in *Resource*, 1976, 6(2).

HPTP had at least 10% of the registrants visit its exhibit with 176 asking for more information. To insure that mostly interested individuals were identified, the address listing was kept away from the front of the exhibit and inquirers had to request to put their addresses on file. All visitors were interested in the date of publication for the modules. Their second question generally dealt with the content and arrangement of the modules. This frequently led into lengthier discussions. A not uncommon subsequent reaction was that visitors often said they were not involved with secondary level students or personnel. Finally, a visitor might inquire about further information. Among these were four people interested in and possibly able to meet HPTP standards for field testing. Another individual had coordinated a curriculum project and wanted to exchange information. In short, a considerable variety of people visited HPTP's exhibit, and an obviously smaller group explicitly expressed interest in using the materials.

The other branch of activity aimed at dissemination was arranging for publishing. This was a crucial test of the work Project members had done because it would determine how long and how far HPTP materials would be dispersed. Notification of the Project's program and materials was sent to 53 publishers. Of these, 9 expressed interest in reviewing the materials. Of these, one publisher said he was interested in the materials for publication. Nevertheless, a period intervened where no communication existed. At the end of this period, this publisher declined to pursue contractual negotiations. Finally, a publisher representative, who had been away from his office during this period, was contacted for the first time. He expressed interest and asked to visit the Project. During the ensuing discussion in April, it became clear that the field-test modules were too cumbersome to publish. Further, the representative confirmed what a number of appraisal and field test writers had stated: the modules did not present the most readable format. Finally, he stated that he was not interested in publishing other publishers' materials (the modules contained considerable material for which reprint rights were to be acquired). These three major criticisms led the Project administrators to consider revision of the modules. Such a revision would be extremely late in the life of the Project, but the review of the criticisms, the high estimate of value placed on publication, and the belief that a revised format would be more efficient to produce, publish and use led them to reduce the material into one volume of readings and activities and a companion volume of test items. The resulting material comprises 1,400 typed pages for the first volume and 600 pages for the test volume.

As of the end of the Project, these materials were prepared to go to the publisher's representative for review. Continued negotiations rest with the Project director and the University. Copies of the materials are submitted with this report.

1. Weisenstein, G. Special project: Optimizing habilitation services in the public schools. First Formative Evaluation Report, 1974.
2. Schoonmaker, W.D. and Girard, J.P. Competencies for special educators: A systematic approach to the training of habilitation personnel. Final Planning Report, 1975.
3. Wimmer, D. Special project: Optimizing habilitation services in the public schools. Second Formative Evaluation Report, 1975.
4. Clark, G.M. Prevocational evaluation in secondary work/study programs.
5. Notley, M. and Girard, J.P. Introducing the HPTP teacher training modules.
6. Clark, G.M. (Ed.) An in-progress outline of the 29 HPTP modules.

TABLE 6. CHRONOLOGICAL LISTING OF PROJECT DOCUMENTS DISSEMINATED

DISCUSSION AND RECOMMENDATIONS

Four major tasks were identified early in the Project: content, training, evaluation, and dissemination. In this section the outcomes of action on these tasks will be weighed.

The content model as realized in the lattice has received high praise and a great deal of interest. It is relatively complicated and was hard for many to understand. Thus the praise and interest take on added significance. It seems clear that the lattice was intellectually sound and could be appreciated as such by those able to understand it.

By the very fact of intellectual excellence for the lattice, social excellence (in this case) is relatively low. The lattice is not in outline form, although with some exceptions, it is essentially an outline. Thus it offers little conceptual support to trainers or trainees accustomed to that format. Further it does not explicitly take into account feedback, which is one feature it should have that is not usually found in an outline. Yet this absence of feedback diagrams has confounded people who have studied the lattice. They have mentioned the cyclic nature of tasks and therefore, the nonlinear nature of the skills represented in the lattice. Nevertheless, the lattice had indirect effects which are socially very worthwhile. Specifically, it provided Project staff with a conceptual foundation for the modules. Thus, the social value of the lattice is in its realization in the modules. This realization was only paramorphic, that is, a tranformation of content into training and not a translation.

The content (lattice) requires incorporation of cyclic feedback mechanisms; then the content should be validated through systematic research. These two activities, although difficult and probably beyond the means of HPTP, should improve the intellectual excellence of the lattice and could conceivably improve its social worth. On balance, the content model realized in the lattice is one of two or three top accomplishments of the Project. This rating is based not only on a comparison of this model to other accomplishments of the Project, but also on a comparison of this content to that found in teacher training materials reviewed by the Program on Teaching Effectiveness at Stanford University. That review consistently showed inadequate conceptualization of content resulted in confusing materials being developed. Despite additional work that could be done, the content model, realized in the lattice, is a major accomplishment of HPTP.

The training model tranformed the content model into usable format at the expense of most of the Project's resources. Unfortunately it was not realized early that this was a transformation process. A straight forward example occurs with the Implementation Module. In the training model, Figure 3, p. 57, this module, I-1, is a practicum at the end of the whole sequence. In the content model, Figure 2, p. 53, it is the third of four functions which a secondary special educator fulfills, i.e., as implementor of a program for exceptional students. The order in which this element appears points to the often significant difference between the use of skills and the order of training to learn those skills. This misunderstanding of the difference between use and training meant resources were used which might have been available for other uses. Specifically, dissemination efforts might have been begun earlier and more personnel devoted to evaluation. There was an overbalance of resources expended on production of the training model.

Part of the outcome of this inbalance was that dissemination, overshadowed at the earlier stages, came to have disproportionate weight at the end of the Project. Other reasons were also crucial in changing the format of the modules; however, the result was that the modules field tested were not the modules prepared for publication. Their format differed. Thus the data presented in this report may not speak to the condition of the modules as completed. Formative data were used in the terminal modification. Nevertheless, summative evaluation on the modules remains to be done. On the other hand, the sequence of modules is complete and will be available for dissemination.

Two recommendations seem appropriate. First, dissemination criteria in the form of publication requirements should be part of the formative evaluation model. This could add a fourth review step which preferably would be carried on outside the Project and in parallel with the development of each module. Secondly, a psychology and sociology of training ought to be incorporated into the modules in an explicit way. The psychology or psychologies would provide a framework which would be clear to users and define Project expectations for the mode and method of learning to be used. The sociology would set up the social contract between the Project and the trainers. The Project encouraged trainers to use a contract form of interaction with trainees; however, this was not clearly defined judging by Instructor Evaluation data. By that very fact, one can see that the conventions established were established in a weak way. Specific statements made during a number of discussions make it clear that incorporating a psychology ar sociology of instruction and field testing would have been difficult. So also would have been the dissemination recommendation. The outcome of these recommendation, nevertheless, would be a more rounded final product, the modules.

The evaluation model implemented was within the resources available in the Project. By its use for formatively gathering data decision making was supported and made empirical. The Formative Evaluation Model represented in Figures 4 - 6 requires only two or three staff people to set up and to collect and analyze data. These people would be the secretary, a field test coordinator, and a data analyst who conceivably could be the coordinator. Such minimal staffing assumes multiple roles for those involved. In fact, HPTP had four to six people who assumed single roles and were involved at various times in this task. These people also worked at other tasks in the Project which were outside evaluation. The point is, that no greater commitment of personnel nor of other resources could have been made to evaluation. Considerable data were collected and analyzed under this arrangement, including achievement data which were developed from tests written by Project staff writing the modules. These data were recycled, as illustrated in Figure 6, for revisions in the training model. In short, the formative evaluation model did what was required of it.

A number of additions to the evaluation model would have been desirable nevertheless. Intellectual data were weak, e.g. reliability and validity. Social data were collected, but some were not systematic, e.g., publishability and generalizability (adaptability).

As noted earlier, reliability and validity may have been out of reach of the Project, not only in terms of resources but in terms of technology. It would have taken another year to institute the methodology for developing parallel criterion or domain referenced tests. By using this methodology, one assumes that the material being tested consists of a network of concepts; thus, dependencies among test items exist. Further, one assumes that testees learn or do not learn elements of that network, but because of the interrelations between elements, they may be able to guess elements they do not know. These assumptions are contrary to those involved in classical reliability and validity. Thus considerable classical work must be proven appropriate to this area of testing or supplanted. HPTP could have tried some of this new methodology if it had been begun two years before the Project ended. This would have given enough time to gear up.

Publishability and generalizability (adaptability) of Project materials were difficult points for the Project. In retrospect, these criteria were more forced on than planned by the Project. Thus in retrospect, it would have been desirable to have more planning in respect to these two. Whether Project resources could have been organized towards these ends is moot.

Overall, the evaluation model worked within the limits imposed by resources. Nevertheless, some things might have been added that are lacking and would be desirable. To provide them, early preparation of the design to account for the above intellectual and social criteria would have been necessary.

One final note on evaluation seems appropriate. Considerable interest in evaluation has been focused on causality. That is, did, for example, MPTP materials *cause* achievement or attitude "improvement" in teacher trainees? Research and evaluation designs have been and can be used to try to answer such questions. Scriven (1974) advocates this in a number of ways. But Stayrook and Corno (1976) present data indicating that at least an interactionist viewpoint is necessary. In short, instructional causality is at least modified by the individual who is being "caused" to learn or act. Finally, McKenzie (1975) presents theoretical arguments, based on research data, which point toward rule-governed actions and away from caused ones. This trend of thinking is manifest in a number of educational research areas, most particularly, and of importance here, in the area of testing. As touched on earlier, domains are being conceived of as being networks of concepts produced by and accessible through rules. In short, the emphasis on causality has proven less useful in this evaluation area of social science than has the conceptualization of rule-governed behavior. These conceptualizations are new enough to have been introduced only in the last year of the Project through the use of attribution analysis of field test data.

The dissemination model was a tested format. As such, when put into operation, it worked smoothly. The outcomes of publicizing were mixed but did contribute to publication and formative needs. Mediation contributed greatly to the effectiveness of the presentations and exhibits. Publicizing was a major accomplishment of the Project.

Publication of the materials remains incomplete at the end of the Project. Earlier contact with a publisher would have been beneficial to the training model, the evaluation model and the dissemination model. This is true for the dissemination model because the publishing arrangements would have been completed early and because marketability would have been assessed. The risk would have been that the other models might have been unduly influenced by salability criteria. Nevertheless, some experimentations might have been possible with incorporating the desired content with a salable format. As it is, content and its preparation have dictated the result. Still, publication activities appear to have worked well, with Project staff partially overcoming difficulties feeding to this area of work from other areas.

On balance, the dissemination model worked effectively, but it remains incomplete as to the desired outcomes. No guarantee could be made that, had publication activities been begun earlier, publication would have resulted. Nevertheless, one could know one way or another at this point if they had been begun earlier.

The Project has made a solid contribution to the field of training for secondary special educators. This includes, most particularly, the conceptual work embodied in the lattice. It also includes the training curriculum realized in the modules. Additional work is to be recommended in meeting intellectual and some social criteria based on a review of the Project's efforts and products. These criteria, it should be noted, probably were out of reach of Project resources available but are customarily expected by the rules governing the kind of work undertaken by HPTP. It is noteworthy, at the same time, that such criteria were not explicitly part of Project tasks. It appears that HPTP accomplished what it promised and tried to, but could not, do all it was tacitly obligated to do. Summatively, HPTP has more work to do, yet has accomplished noteworthy work in what it did do.

APPENDIX A

ADJUNCT RESEARCH

COX-VARGAS PPDI ANALYSIS OF SOME ITEMS FROM THE WORLD OF WORK

Lynne Prouty

The movement towards a competency-based system of teacher training can be seen as one attempt to meet a number of concerns facing the field of education today. The related demands for relevancy of educational programs and accountability on all levels can be seen as the fundamental incentives which ultimately led to the competency or performance based teacher education (C/PBTE) approach. Other factors which influenced the shape and direction of this movement include the perpetual search to improve the quality of education, the administrative need to focus on cost-effectiveness issues, widespread efforts to emphasize individualized education, and the push to spread systems management into the educational realm. (Schmieder, 1973; Houston, 1974; Rosner & Kay, 1974.)

A number of definitions or descriptions of competency-based educational programs have been reported in the last few years. However, Elam (1971) has suggested that C/PBTE programs have these essential elements in common:

(1) Competencies (knowledge, skills, behaviors) to be demonstrated the student are

Derived from explicit conceptions of teacher roles;

Stated so as to make possible assessment of a student's behavior in relation to specific competencies;

and

Made public in advance.

(2) Criteria to be employed in assessing competencies are

Based upon, and in harmony with, specified competencies;

Explicit in stating expected levels of mastery under specified conditions;

and

Made public in advance.

(3) Assessment of the student's competency

Uses his performance as the primary source of evidence;

Takes into account evidence of the student's knowledge relevant to planning for, analyzing, interpreting, or evaluating situations of behaviors;

and

Strives for objectivity.

(4) The student's rate of progress through the program is determined by demonstrated competency rather than by time or course completion.

(5) The instructional program is intended to facilitate the development and evaluation of the student's achievement of competencies specified.

Several C/PBTE projects have been organized at various universities and have developed large bodies of materials. Three widely recognized examples are reviewed very briefly below in respect to their assessment. Then HPTP's efforts are discussed.

W. Robert Houston at the University of Houston in Texas developed a comprehensive competency based program for elementary and secondary teacher trainees in general education. The competencies were generic, therefore, useful for this broad range of teachers. They were classified into four groups: (1) knowledge, (2) performance, (3) consequence (student achievement), and (4) exploratory (experiential activities) competencies. Measurement was most rigorous for the first and third classes of competencies where relatively standard methods of teaching and observation could be used. Criterion behavior was arrived at by surveying authoritative opinion. The Houston program found measurement and assessment of criterion behavior to be difficult problems.

Illinois State University has provided a modular instructional sequence for secondary teachers which is a substitute for the standard course sequence. The topics seem relevant but uncoordinated. The activities and tests while conceptually appropriate are topic specific and perhaps too current to be of long lasting value. This raises the dilemma between general content which is difficult to objectively assess and specific and contemporary content which, though currently relevant and assessible, may not prove to be of general survival value. Measurement issues seem to have been given secondary consideration in this effort to provide a sequence of study which was more objective than previous curricula.

At the School of Education at Weber State College, a third body of individualized, performance-based teacher education materials, "Wilkits," were devised. They contain objectives, some readings, recommended reading materials and outlines for evaluation. The performance aspect is somewhat unrealized partly because reading is a major foundation for the kits and partly because behavior or performance is defined at a relatively general level. The tasks, often including writing, do however include practice at skills. Evaluation is primarily the responsibility of the instructor. Evaluation instruments are minimal, though standards are provided.

The assessment portions of these C/PBTE programs appear less objective than one might expect. Performance certainly is observable, but, as we shall see in reviewing one HPTP test, the measurement of that performance using test items is hazardous.

The Habilitation Personnel Training Project (HPTP) is a federally funded project within the Department of Special Education at the

University of Kansas, which also developed a C/PBTE program. This project was developed for the purpose of designing an instructional training program for the preparation of secondary special educators; specifically, classroom teachers and work/study coordinators.

The project utilizes a C/PBTE approach realized in a modular format with appended tests of performance. This format satisfies specifications set by Cooper and Weber (1973) as desirable for C/PBTE instructional modules. In short, the modules were prepared to realize recommended practice.

One of the problems, referred to above, in the development of any C/PBTE program lies in the construction of assessment instruments that can be used to measure a student's level of performance with respect to criterial objectives, competencies, and performance (Schmieder, 1973; Andrews, 1974, Dodl, 1973.) The focus on test instruments can be found in the earliest writing in this area.

Glaser, in his landmark article of 1973, began the discussion of criterion-referenced measurement. He considered the task of evaluating student performance to be one of " . . . measurement of subject matter proficiency as it may be defined by recognized subject matter scholars." (p. 519) In his view,

> . . . criterion levels can be established at any point in instruction where it is necessary to obtain information as to the adequacy of an individual's performance. . .

> (and) . . . specific behavior implied at each level of proficiency can be identified and used to describe the specific tasks a student must be capable of performing before he achieves one of these knowledge levels. It is in this sense that measures of proficiency can be criterion-referenced. (p. 519)

From this beginning, the literature in the field mushroomed. Since Glaser's 1963 article, numerous related definitions and terminologies have appeared. "Content-standard," universe-defined," "objective-based," "mastery tests," "domain-referenced," "criterion-referenced differential assessment device (CRDAD)", etc. all have similar, though in no way interchangeable, definitions (Millman, 1974).

For the purposes of HPTP, a criterion-referenced test (CRT) is defined as a measure which " . . . used to ascertain an individual's status with respect to some criterion, i.e., performance standard" (Popham, and Husek, 1971, p. 20).

Criterion-referenced tests have been generally accepted as the preferred form of measurement for use with C/PBTE programs (Coulson and Cogswell, 1965; Glaser and Cox, 1968). However, the development of a valid CRT is quite difficult. There are some vexing problems that manifest themselves in CRT construction. Harris (1974) states that "A problem . . . for tests that are designed to be integrated into a specific instructional system is that of how much "instructional bias' is to be either allowed or demanded" p. 85. Millman (1974) shows that the generation, selection, and statistical analyses of test items and the establishment of passing standards (or criterion levels) are two general problem areas integral with the construction of CRT'S. These two problems were basic to those encountered by the HPTP, and it was out of a need for possible solutions that this study arose.

In the early stages of the project, it was decided that alternate forms of the same test should be used as pre and post measures, rather than using an identical test for both purposes. Establishing the reliability and validity of instruments constructed this way and determining criterion levels for mastery were problems left to be solved later.

The purpose of this study was to assist in the solution of this variety of problems for one HPTP module, *The World of Work*. By administering items from this module to three groups of subjects---trainees entering the module (pre-test), trainees exiting from the module (post-test), and teachers in the field giving instruction related to the world of work but who hadn't been trained with the module (experienced),--data were developed on the items. Three outcome indexes were to be obtained by comparing the data for each group with the other groups. On examining these indexes, the ability of items to discriminate among the three groups as well as the utility of each distractor was to become clear. Through analysis of the indexes sensitivity of the test to the needs of teachers in the field, or mere sensitivity to module instruction, would be determined. Finally, these indexes were expected to contribute towards developing a methodology for setting valid criterion levels for our tests. Until criterion levels were established, the *World of Work* instrument could in no way be considered truly criterion-referenced nor could our other modules.

While item analysis has been recently supported for use with CRT's, (Haladyna and Roid, 1976), the question of which procedures are best leads to discussion. Cox (1971) suggests that traditional approaches " . . . are either not applicable or are difficult to adapt to the criterion-referenced framework . . ." and that alternative procedures need to be explored.

Popham and Husek (1971) make a case for modifying reliance on item discrimination indexes for CRT's from that of the usual norm-referenced test (NRT) usage to considerations of CRT requirements. For instance, an item which does not discriminate among subjects at all would be eliminated in an NRT. However, if the item were deemed fundamental to the objective and criterion in a CRT, it would be retained. Furthermore, such an item, if it had a positively discriminating but a low value, might point to weaknesses in instruction. A negatively discriminating CRT item, however, would be interpreted the same way as it would be for an NRT; that is, there may be a number of explanations for the index, and one should analyze the item carefully for flaws.

Nevertheless, Haladyna and Roid (1976) supported four out of eleven item discrimination indexes, used on a CR test, to be the best measures of the quality of CRT tests and of the sensitivity of the items to instructional content. They concluded that " . . . the (Cox-Vargas) PPDI (pre-to-post difficulty index) comes closest to measuring item sensitivity in a direct and simple manner which is conceptually satisfying when one considers the nature of instructional sensitivity at the test level, i.e., the difference between pre-and post-instruction means" (p. 20).

The PPDI is the result of the Cox and Vargas (1966) report on two discrimination indices. Items were administered as a pre- and as a post-test. The first index applied was the traditional upper-group minus lower group on the post test, which gave information as to how well each item discriminated between students. The second applied was the PPDI index, which used both pre- and post-test data. By subtracting the percent of students who passed the item on the pre-test from the percent who passed the item on the post test, the PPDI resulted. This index provided information as to how well each item discriminated between pre- and post-test groups, that is, between uninstructed and instructed groups. When Cox and Vargas compared the two indexes, they found that some items which proved to be highly useful for discriminating pre- and post-test groups with their novel analysis would have had to have been discarded under the traditional upper minus lower analysis procedure. Some items failed to discriminate among individuals, but did discriminate between instructed and uninstructed trainees. Haladyna and Roid (1976) supported the usefulness of the Cox-Vargas PPDI procedure for use with criterion-referenced tests.

METHODOLOGY

Each HPTP module consists of three components: a trainee's booklet, an instructor's booklet, and a readings booklet. The recommended sequence of movement through an HPTP module is typical of most C/PBTE modular formats. The trainee is invited to read the rationale and objectives for the module. He is then given a diagnostic pre-test which provides him/her with an opportunity to demonstrate prior mastery of any objectives. At that point, the trainee is directed to recommend activities for any objectives which he/she has not met as indicated by the pre-test. A post-test is later administered to evaluate the trainee's achievement of these objectives. The trainee is allowed to repeat the post-test until proficiency is reached.

The objectives for each HPTP module were developed with the intent of reflecting knowledge directly applicable in secondary special education settings. Assessment instruments were prepared to serve as pre and post-tests, and were to indicate a trainee's mastery of behavior referred to by each objective, rather than to merely indicate his/her knowledge about information from the recommended readings in the module.

The test items themselves were constructed by various members of the HPTP staff. These items were usually objective in nature, with some situational problems presented for the trainee to solve. The items were intended to represent a variety of taxonomic levels of the cognitive domain as established by Bloom, et. al. (1956)

In order to conduct an analysis of the content sensitivity of these tests, uninstructed (pre-test) instructed (post-test) and experienced, subjects were gathered. Uninstructed subjects were operationally defined as students enrolled in a special education program at either the undergraduate or graduate level, and who had not yet studied, given instruction in, or organized work-related activities for the handicapped. Either these subjects were given the instrument as a pre-test in courses using the *World of Work* module, or they were voluntary participants from other special education courses who received the half tests discussed below.

Instructed subjects were those who successfully completed the *World of Work* module in the course of their university studies, but who had no content-related teaching experience.

To obtain experienced subjects, professors of special education at three universities were asked to provide names of teachers whom they considered experienced in giving work-related activities to secondary special education students but whom they knew had no contact with the module. Those individuals named were contacted by mail. The purpose of the study was explained, their cooperation was solicited, and a copy of a modified test was enclosed. In addition, a Demographic Data Form, Subappendix A, was enclosed, which provided information as to the nature of each of the participating subjects' work-related special education experience. This was done in order to guarantee that each "experienced" subject met the required specifications. While their participation was voluntary, a small sum of money ($2.50) was offered to each experienced person who completed and returned the test according to instructions.

It was estimated in advance that the entire test would take well over an hour to complete. Because it was feared that many potential experienced subjects would not be willing to contribute so much time, participating members of this group were given modified half-tests. In order to matrix sample the test and to control dependencies among items put on the half-tests, each half-test consisted of items randomly selected from the total item pool. Thus, no two subjects received exactly the same test.

No time limit was set for taking the test, but it was suggested that it would take approximately thirty minutes to complete.

Of the teachers contacted, 26 or 65% participated.

Because of the random selection of items for each test completed by experienced subjects and because of unequal numbers of subjects among the three groups, there were many items for which the sizes of the groups differed. In these cases, subjects' responses from the larger groups were eliminated by random rejection until all groups were of equal number.

Any item answered by two groups, of at least ten subjects each, was analyzed by the Cox-Vargas PPDI method. When an item was answered in sufficient numbers by all three groups, the Cox-Vargas was used in three combinations: uninstructed (pre) % from instructed (post) %, pre % from experienced %, and post % from experienced %.

In addition to the PPDI, which uses only the percentages of subjects from each group responding to the correct alternative, a distractor analysis was conducted for each item, which consisted of computing the percent of subjects in each group marking each incorrect response.

RESULTS AND CONCLUSIONS

The data from the Cox-Vargas item analysis indicates that the test did not achieve its intended purpose. Nevertheless, several points of interest resulted from the analysis and are discussed below.

There were 21 objective items in which sufficient numbers of subjects, from at least two of the three groups, responded to compute a PPDI. The results are presented in Table 1.

	col. 1	col. 2	col.3	col. 4	col. 5
TYPE OF PPDI	NUMBER OF ITEMS	%*	NUMBER OF POSITIVE PPDI	% OF COL. 1	%*
1. Pre from Post (1x2)	14	67	11	79	52
2. Post from experienced (2x3)	12	57	2	17	10
3. Pre from experienced (1x3)	18	86	3	17	14

*21 items overall had one or more PPDI's.

TABLE 1

Summary of items with a PPDI
(Pre to Post Difficulty Index)

On the positive side, the post-test group performed better than the pre-test group on 52% of the items (11/21), and 79% were better than the pre-test group on items specifically analyzed with a pre from post PPDI. This indicates that those items are sensitive to instruction. If this were the goal, the test items would be relatively good. In contrast, the post-test group was equal to or better than the experienced group on 90% of the items in which comparisons could be made. The test items contained only five on which the experienced group performed as well as or better than both of the other two groups. On two items (15 and 16), experienced were better than post; on three other items (2, 3 and 10), experienced were better than pre. It is obvious that the test was sensitive to instruction of the module, but not to the knowledge and abilities of experienced teachers.

In other words, these data show serious weaknesses in the test. The pre-test group performed equal to or better than the experienced group on 71% ((18-3)/21) of the items in which comparisons could be made. Thus the experienced group's results cannot be used for making decisions as to criterion-levels, as had originally been suggested. Furthermore, the pre-test group performed equal to or better than the post-test group on three items which certainly must be revised or eliminated.

It is surprising how serious this problem is; that is, how many of the experienced group's PPDI are equal to or lower than those from the pre-test group: 83% of the 18 items common to both. There are a number of possible explanations for this phenomenon. (1) Perhaps it was because the experienced group received their tests and instructions by mail, while the other two groups were in a classroom setting. (2) Some errors in collating of the mailed tests resulted in unclear directions which also could have affected outcomes. (3) It is possible that the pre- and post subjects gained an advantage in terms of sequential dependencies among items that the random selection of items for the experienced group discouraged. (4) Perhaps those people in the field are no longer as proficient at taking tests as are those who are still in a university environment. (5) It is quite likely that the pre- and post test groups were more motivated to do well since their performance affected their course grades. (6) Many practicing secondary special education teachers were not trained specifically for secondary education and therefore may not have taken work-related coursework. (7) Finally, it is possible that the content deemed important for this module is not consistent with the knowledge actually needed by classroom teachers.

The distractor analysis suggests that mistakes in test construction were separable from the discriminability of the items discussed above. Seven of the 21 items contained at least one distractor that was unanimously rejected. Item 5 was performed correctly by everyone, so the three distractors were of no use. Several more used a list of ten distractors, many that were of dubious utility for some items. These were "matching" items, so it is not surprising to find that improbable choices existed. This format warrants review and probable rejection when the test is revised.

One specific weakness in the study may have contributed to these outcomes: The relatively small numbers of subjects in each group, which affects the conclusions drawn. The results may not even be generalizable to the full number of *World of Work* items. Certainly, it would have been advantageous to have been able to compute a PPDI for all three groups on every item under study.

In summary, it appears that the use of the *World of Work* test in its present form is not justified as a criterion-referenced test. Reconstruction of the test should perhaps begin with a test "blueprint." As described by Thorndike and Hagen (1969), such a blueprint requires that the constructor " . . . state explicitly the purpose for which the test is to be constructed [not done by HPTP], define in operational terms the objectives that are to be measured by the test [done], determine the content to be covered by the test [done] and decide upon the relative importance of each of the objectives and each area of content to be reflected in the test [not specifically done]." (p. 34) As can be seen by the bracketed notes, two major points recommended for the preparation of tests were inadequately done in the preparation of this test.

Yet to be decided is the question of criterion-levels. It was hoped that data from the experienced subjects' scores could be used to develop a paradigm for setting a criterion. Such an outcome is not possible given the existing data. Further, the outcomes from these data suggest unsuspected difficulties with the paradigm in any event. This study only seems to provide evidence which complicates the use of experienced testees as criterion setters.

The results of this study are shaky evidence of problems alluded to by other investigators. The conclusions are that the *World of Work* test items are instructionally sensitive but possibly invalid with respect to experienced teachers' knowledge and skills. A replication with adequate sample size is needed.

References

Andrews, T. E., What we know and what we don't know. In W. R. Houston, *Exploring competency based education*. Berkeley, CA: McCutchan Publishing Corp., 1974.

Bloom, B. S. (Ed.), Englehart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. *A taxonomy of educational objectives: Handbook I, the cognitive domain*. New York: Longmans, Green, Co., 1956.

Cooper, J. M. and Weber, W. A. A competency based systems approach to teacher education. In D. W. Anderson; J. M. Cooper; M. V. DeVault; G. E. Dickson; C. E. Johnson; and W. A. Weber, *Competency based teacher education*. Berkeley, CA: McCutchan Publishing Corp., 1973.

Coulson, J. E., and Cogswell, J. F. Effects of individualized instruction on testing. *Journal of Educational Measurement*, 1965, 2(1), 59-64.

Cox, R. C. Evaluative aspects of criterion-referenced measures. In W. J. Popham (Ed.), *Criterion-referenced measurement: An introduction*. Englewood Cliffs, N. J.: Educational Technology Publications, 1971.

Cox, R. C. and Vargas, J. S. A comparison of item selection techniques for norm-referenced and criterion-referenced tests. Paper read at the annual meeting of the National Council on Measurement in Education, Chicago, February, 1966, as reported by R. C. Cox, in W. J. Popham (Ed.), *Criterion-referenced measurement: An introduction*. Englewood Cliffs, N. J.: Educational Technology Publications, 1971.

Dodl, N. R. Program evaluation and student assessment. In D. W. Anderson; J. M. Cooper; M. V. DeVault; G. E. Dickson; C. E. Johnson; and W. A. Weber, *Competency based teacher education*. Berkeley, CA: McCutchan Publishing Corp., 1973.

Elam, S. *Performance-based teacher education: What is the state of the art?* Washington, D. C., American Association of Colleges for Teacher Education, 1971.

Glaser, R. Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 1963, 18(8), 519-521.

Glaser, R. and Cox, R. C. Criterion-referenced testing for the measurement of educational outcomes. In R. A. Weisgerber (Ed.), *Instructional process and media innovation*. Chicago: Rand-McNally, 1968.

Haladyna, T. M. and Roid, G. H. The quality of domain-referenced test items. Paper read at the Annual Meeting of the American Educational Research Association, San Francisco, 1976.

Harris, C. W. Problems of objective-based measurement. In C. W. Harris, M. C. Alkin, and W. J. Popham (Eds.), *Problems in criterion-referenced measurement*. Los Angeles: Center for the Study of Evaluation, 1974.

Houston, W. R. (Ed.) *Exploring competency-based education*. Berkeley, CA: McCutchan Publishing Corp., 1974.

Millman, J. Criterion-referenced measurement. In W. J. Popham (Ed.), *Evaluation in education: Current applications*. Berkeley, CA: McCutchan Publishing Corp., 1974.

Popham, W. J. and Husek, T. R. Implications of criterion-referenced measurement. In W. J. Popham (Ed.), *Criterion-referenced measurement: An introduction*. Englewood Cliffs, N. J.: Educational Technology Publications, 1971.

Rosner, B. and Kay, P. M. Will the promise of C/PBTE be fulfilled? *Phi Delta Kappan*, January, 1974, 290-295.

Schmieder, A. A. *Competency-based education: The state of the scene.* Washington, D. C.: American Association of Colleges for Teacher Education, ERIC Clearinghouse on Teacher Education, and National Center for the Improvement of Educational Systems, 1973.

Thorndike, R. and Hagen, E. *Measurement and evaluation in psychology and education*. New York: John Wiley and Sons, 1969.

SUBAPPENDIX A

1. DEMOGRAPHIC DATA FORM
2. LETTER TO EXPERIENCED SUBJECTS

Demographic Data Form

1. Subject #__________

2. Classification (at present):

   0-15 Grad Hours_____(1) 16-30 Grad Hours_____(2) 31-45 Grad Hours_____(3) Over 45 Grad Hours_____(4)

3. Experience in Special Education:

   (1) No Teaching Experience_____
   (2) Student Teaching_____weeks
   (3) Elementary Special Education Teaching_____years
   (4) Secondary Special Education Teaching_____years
   (5) Work-study Coordinator_____years
   (6) Special Education Administrator_____years
   (7) Summer camps_____weeks
   (8) Institutional Work_____years
   (9) YMCA or YWCA, or other recreational work_____weeks
   (10) Other_____

4. Have you studied, given instruction in or organized work related activities for the handicapped?_____Yes_____No

Letter To Experienced Subjects

Dear

For the past four years, members of the Habilitation Personnel Training Project, under the directorship of Dr. Gary Clark, have been developing a competency-based training program for preparing teachers of secondary special education students. We have identified competencies needed by secondary special education personnel, and have written instructional modules designed to teach those competencies. These modules are currently being field tested at colleges and universities throughout the United States.

We are now collecting data which will enable us to validate the tests within the modules. To do this, we are asking a number of secondary special educators--like yourself--to take a sample test. We hope this information will make our modules reflect more accurately the real needs of teachers. I am sending you a portion of the pre-test from our module on "The World of Work," which deals with the capabilities needed by the secondary special education student for entering and successfully participating in a viable occupational area.

I am sure that you are very busy with your school activities. However, the test should take less than thirty minutes of your time, and as a token of our appreciation, we are prepared to pay you $2.50 for completing and returning the enclosed materials. Won't you complete it now while it is fresh in your mind.

Please be assured that your response will be held in confidence. Your name will not appear in the study or be used for anything other than the purposes stated above. We have number coded the test copies only to be able to keep track of those who have responded. You may withdraw at any time, even if you have returned the test to us.

Thank you for your cooperation. The information you provide us will surely enable us to do a better job of preparing future special educators. If you wish to read our final report, let me know.

Sincerely,

Lynne Prouty
Research Writer
309 Carruth-O'Leary
University of Kansas
Lawrence, Kansas 66045

LP/sh
Enclosures

P.S. These materials must be postmarked no later than in order for you to receive payment.

SUBAPPENDIX B

ITEM ANALYSIS

SAMPLE

Item Analysis Sample

TYPE OF PPDI: A: Pre from Post *(ie., 2-1)*
B: Post from Experienced *(ie., 3-2)*
C: Pre from Experienced *(ie., 3-1)*

QUESTIONS:

1. On the job, any individual must not only have to learn his job skills, but must (select the best answer):
   a. learn what behaviors are expected of him on the job.
   b. be able to avoid other activities on the job which aren't job skills
   c. have the ability to get along well with his associates.
   d. set goals for himself to complete on the job.

TEST GROUP	no. of sub.	RESPONSES TO QUESTION 1.				
		A *	B	C	D	OMIT
1. PRE	11	8= %73	2= %18	1= %9	0	0
2. POST	11	11=%100	0	0	0	0
3. EXPERIENCED	11	5= %45	0	5= %45	1= %9	0
A. 2-1		%27				
B. 3-2		-%55				
C. 3-1		-%28				

B. 3-2			
C. 3-1		0	

QUESTIONS 8-15

MATCHING: Enter the letter of a competency listed on the left in each blank on the right for item s) listed below:

Competencies:
A. Ability to select a vocation.
B. Ability to select a particular job opening for which to apply
C. Ability to apply for the job.
D. Ability to hold a job.

Which competency does one have when he/she knows:

_____ 8. which jobs have extremely hot or cold working conditions.

TEST GROUP	no. of sub.	RESPONSES TO QUESTION 8.				
		A*	B	C	D	OMIT
1. PRE	12	7= %58	4= %33	0	1= %8	0
2. POST	1	1	0	0	0	
3. EXPERIENCED	12	5= %42	5= %42	0	2= %17	0

A. 2-1		
B. 3-2		
C. 3-1	-%16	

_____ 9. how to wait before accepting a job without losing a chance at the job.

TEST GROUP	no. of sub.	RESPONSES TO QUESTION 9.				
		A	B	C*	D	OMIT
1. PRE	11	2= %18	2= %18	4= %36	3= %27	0
2. POST	11	1= %9	1= %9	7= %64	2= %18	0
3. EXPERIENCED	11	0	1= %9	7= %64	3= %27	0
A. 2-1				%28		
B. 3-2				0		
C. 3-1				%28		

_____10. how to use sources of job information (e.g. newspapers.)

TEST GROUP	no. of sub.	RESPONSES TO QUESTION 10.				
		A	B*	C	D	OMIT
1. PRE	15	5= %33	7= %47	3= %20	0	0
2. POST	1	0	1	0	0	
3. EXPERIENCED	15	1= %7	7= %47	7= %47	0	0

APPENDIX B

1. Instructor Evaluation
2. Module Evaluation

UNIVERSITY OF KANSAS
DEPARTMENT OF SPECIAL EDUCATION

Competency-Based Teacher Model

INSTRUCTOR EVALUATION

Module Name____

Instructor's Name__________________________________

Institution____

[4] Did you find the instructions in the instructor's manual adequate?_____If not, please suggest improvements.

[5] Did students seem to require much clarification of the instructions, etc., in the trainee's booklet?_____If so, please suggest possible improvements.

[3] About how often did you confer with each trainee about the Module?_____

[2] Were you able to allow trainees essentially complete freedom to complete the Module as they desired?_____If not, please explain.

[6] Were the suggested resource materials adequate?_____If, not, do you have additional suggestions?

[10] Were the suggested resource materials easily obtained?_____

[9] Did you find it possible to keep adequate notes on the instructor-trainee conferences?_____If not, please explain.

[12] Did you find the pretest, answer key, and accompanying placement suggestions satisfactory?_____If not, please explain.

[7] Did you find the post-test and accompanying answer key satisfactory?_____If not, please explain.

[8] Was the level of presentation appropriate for most of your students?_____If not, please explain.

[11] Was the content appropriately sequenced?_____If not, please explain.

[13] How were the HPTP produced articles and audio-visual materials received by your students?

[14] Was the module format appealing?_____If so, please explain.

[15] Did the module meet your instructional needs?_____Please explain.

[16] Did most students consult with you as suggested in the module instructions?_____

[1] What is your overall evaluation of this Module?

UNIVERSITY OF KANSAS
DEPARTMENT OF SPECIAL EDUCATION

HABILITATION PERSONNEL TRAINING PROGRAM

Competency-Based Teacher Model

MODULE EVALUATION

Student ID No.______________ School___________________ Date__________

Module number or Title____

Estimated number of hours to complete this module___________________

Please circle the one response which is most indicative of your evaluation of the module. Use the following key:

SA - Strongly Agree
A - Agree
N - Neither
D - Disagree
SD - Strongly Disagree

Your Personal Response

1.	In view of the effort I put into the module, I was satisfied with what I gained.	SA	A	N	D	SD
2.	This module provides information which tells me how to teach the handicapped.	SA	A	N	D	SD
3.	To meet the objectives, I memorized information that did not seem very important.	SA	A	N	D	SD
4.	The module stimulated my thinking about this area of special education.	SA	A	N	D	SD
5.	I worked at my own pace in the module.	SA	A	N	D	SD
6.	The amount I learned was commensurate with the effort required to satisfactorily complete the module.	SA	A	N	D	SD
7.	I like the competency based format in which objectives, activities, and assessment are used.	SA	A	N	D	SD
8.	The topic of the module is especially interesting to me.	SA	A	[illegible]	[illegible]	SD

9. Content in this module will help me in my work. SA A N D SD

10. The competency based format, as used here, made me confident of my knowledge. SA A N D SD

11. I changed by attitudes as a result of completing this module. SA A N D SD

12. Skills presented in this module were new to me. SA A N D SD

13. As I worked on the module, I had the opportunity to discuss interesting issues with the instructor. SA A N D SD

14. The module was a worthwhile learning experience. SA A N D SD

15. After reading the objectives, I felt I would be able to achieve them. SA A N D SD

16. The module made sense to me. SA A N D SD

17. I would take another course offered in this manner. SA A N D SD

18. The module format was appealing to me. SA A N D SD

The Objectives

19. The objectives are about real problems. SA A N D SD

20. The most important objective in this module is ______________________

21. The module should provide a more comprehensive explanation of what the learner is expected to do. SA A N D SD

22. Some indication should be given of the amount of time to complete a module. SA A N D SD

23. The objectives match those of work-study coordinators. SA A N D SD

24. It is difficult to determine from the objectives exactly what is expected of the learner. SA A N D SD

25. The objectives match those of the special education teacher. SA A N D SD

26. An instructor should be readily available for consultation whenever necessary. SA A N D SD

27. The focus of the module is clear. SA A N D SD

28. The instructor presented clear, general instructions at the beginning. SA A N D SD

The Content

29. The information presented in the enabling activities is new. SA A N D SD

30. This content is interesting. SA A N D SD

31. The readings were too long. SA A N D SD

32. The suggested activities were needed to make the instruction complete. SA A N D SD

33. The issues which were presented were handled well. SA A N D SD

34. The topics which were discussed covered the important issues. SA A N D SD

35. Another issue of importance is ____________________

36. Emphasis is placed on ways to teach the handicapped. SA A N D SD

37. Interrelationships between concepts which were presented are easy to understand. SA A N D SD

38. I had no difficulty identifying the important portions of the assigned readings. SA A N D SD

39. Concepts were well defined. SA A N D SD

40. The enabling activities provided a sufficient number of instructional alternatives from which to choose. SA A N D SD

41. The activities provide practice at important skills. SA A N D SD

42. The enabling activities were useful in helping me master the objective(s). SA A N D SD

43. Values are treated fairly. SA A N D SD

44. Each student should have his own book of readings, even if he has to pay for it. SA A N D SD

45. Graduate level material should be offered. SA A N D SD

46. There should be more on-site visitations among the enabling activities. SA A N D SD

47. Discussion should more often be an activitiy. SA A N D SD

48. Information in this module is essential for preparing secondary special education teachers. SA A N D SD

49. Examples of described materials would be helpful. SA A N D SD

50. Dress rehearsals of teaching techniques should be an alternate activity. SA A N D SD

51. The objectives could be mastered more efficiently by some other means. SA A N D SD

52. There should be an option for students to propose alternate enabling activities. SA A N D SD

The Assessment Procedures

53. Mastery of the objectives was tested in the manner that I expected. SA A N D SD

54. Being assessed on prespecified objectives makes good sense. SA A N D SD

55. Short-answer and essay would be better than multiple-choice items. SA A N D SD

56. The tests tested what was specified by the objectives. SA A N D SD

57. The tests measure information gained from the readings and activities. SA A N D SD

58. When I had completed the enabling activities I knew I had successfully mastered the objective. SA A N D SD

59. The test format has the right amount of "think" questions in it. SA A N D SD

60. From the wording of some of the post-test items, I was uncertain of what was being asked. SA A N D SD

61. There were too few items on basic knowledge gained from the module. SA A N D SD

62.	Methods other than paper and pencil tests should have been used for evaluating mastery of the stated objectives.	SA	A	N	D	SD
63.	The "think" questions can be answered by referring to the content provided.	SA	A	N	D	SD
64.	My responses to the post-test items provide an accurate representation of what I learned.	SA	A	N	D	SD
65.	The skills training is tested by the assessment procedures.	SA	A	N	D	SD
66.	I knew exactly what standards would be used in judging the quality of my performance.	SA	A	N	D	SD
67.	Issues about values are fairly assessed.	SA	A	N	D	SD
68.	Having the instructor give the tests interfered with self pacing.	SA	A	N	D	SD
69.	The pre-tests helped me prepare for the following competency unit.	SA	A	N	D	SD

The Resources

70.	There were plenty of materials available to complete this module.	SA	A	N	D	SD
71.	The readings are adequate resource materials.	SA	A	N	D	SD
72.	Readings not included in the module are easy to find.	SA	A	N	D	SD
73.	Materials to carry out the enabling activities are available at my school.	SA	A	N	D	SD
74.	No training is necessary to use the materials for the enabling activities.	SA	A	N	D	SD
75.	The enabling activities carried out in a group require a well prepared leader.	SA	A	N	D	SD

PLEASE IDENTIFY ONE ENABLING ACTIVITY WHICH IS ESPECIALLY VALUABLE:

PLEASE IDENTIFY ONE ENABLING ACTIVITY WHICH IS OF MINIMAL VALUE:

ADDITIONAL COMMENTS: (on back)

APPENDIX C

17 Factor Clusters of Items

Factor 1 (26.1% of total variance)	Loadings
10. The competency based format, as used here, made me confident of my knowledge.	.477
38. I had no difficulty identifying the important portions of the assigned readings.	.534
39. Concepts were well defined.	.521
40. The enabling activities provided a sufficient number of instructional alternatives from which to choose.	.497
41. The activities provide practice at important skills.	.487
53. Mastery of the objectives was tested in the manner that I expected.	.632
56. The tests tested what was specified by the objectives.	.759
57. The tests measure information gained from the readings and activities.	.691
58. When I had completed the enabling activities, I knew I had successfully mastered the objective.	.575
59. The test format has the right amount of "think" questions on it.	.552
60. From the wording of some of the post test items, I was uncertain of what was being asked.	-.565
63. The "think" questions can be answered by referring to the content provided.	.566

Factor 2 (4.6% of total variance)	Loadings
1. In view of the effort I put into the module, I was satisfied with what I gained.	.517
4. The module stimulated my thinking about this area of special education.	.656
8. The topic of the module is especially interesting to me.	.787
9. Content in this module will help me in my work.	.742
14. The module was a worthwhile learning experience.	.684
16. The module made sense to me.	.638
18. The module format was appealing to me.	.545
30. This content is interesting.	.784
31. The readings were too long.	-.487
33. The issues which were presented were handled well.	.465
48. Information in this module is essential for preparing secondary special education teachers.	.659

Factor 3 (3.9% of total variance)

27. The focus of the module is clear.

28. The instructor presented clear, general instructions at the beginning.

66. I knew exactly what standards would be used in judging the quality of my performance.

Factor 4 (3.8% of total variance)

2. This module provides information which tells me how to teach the handicapped.

36. Emphasis is placed on ways to teach the handicapped.

46. There should be more on-site visitations among the enabling activities.

Factor 5 (3.2% of total variance)

43. Values are treated fairly.

67. Issues about values are fairly assessed.

Factor 6 (2.9% of total variance)

70. There were plenty of materials available to complete this module.

72. Readings not included in the module are easy to find.

73. Materials to carry out the enabling activities are available at my school.

Factor 7 (2.9% of total variance)

12. Skills presented in this module were new to me.

29. The information presented in the enabling activities is new.

Factor 8 (2.8% of total variance)

15. After reading the objectives, I felt I would be able to achieve them.

54. Being assessed on prespecified objectives makes good sense.

Factor 9

51. The objectives could be mastered more efficiently by some other means.

52. There should be an option for students to propose alternate enabling objectives.

Factor 10

47. Discussion should more often be an activity.

Factor 11

23. The objectives match those of work-study coordinators.

Factor 12

5. I worked at my own pace in the module.

69. The pre-tests helped me prepare for the following competency unit.

Factor 13

49. Examples of described materials would be helpful.

50. Dress rehearsals of teaching techniques should be an alternate activity.

Factor 14

7. I like the competency based format in which objectives, activities, and assessment are used.

32. The suggested activities were needed to make the instruction complete.

Factor 15

74. No training is necessary to use the materials for the enabling activities.

Factor 16

44. Each student should have his own book of readings, even if he has to pay for it.

Factor 17

55. Short-answer and essay would be better than multiple-choice items.

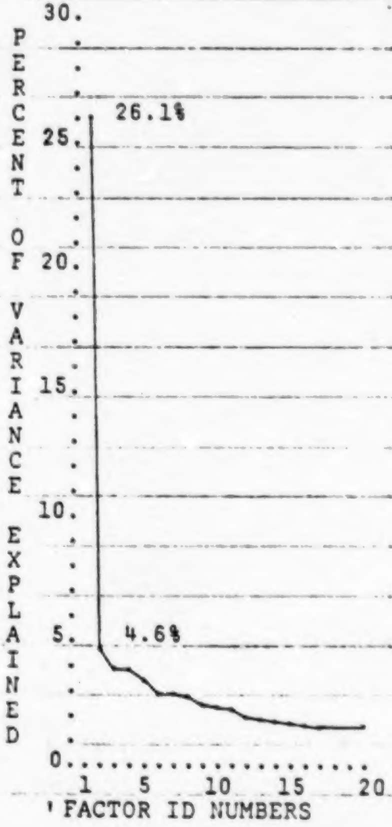
APPENDIX D

Data for 54 Factors

and

"Skree" Analysis

FACTOR	VARIANCE EXPLAINED	PROPORTION OF TOTAL VARIANCE	CUMMULATIVE PROPORTION OF TOTAL VARIANCE
1	17.982	0.261	0.261
2	3.202	0.046	0.307
3	2.694	0.039	0.346
4	2.603	0.038	0.384
5	2.177	0.032	0.415
6	2.032	0.029	0.445
7	1.977	0.029	0.473
8	1.942	0.028	0.502
9	1.574	0.023	0.524
10	1.529	0.022	0.547
11	1.447	0.021	0.568
12	1.407	0.020	0.588
13	1.316	0.019	0.607
14	1.206	0.018	0.624
15	1.145	0.017	0.641
16	1.113	0.016	0.657
17	1.038	0.015	0.672
18	0.998		0.687
19	0.938		0.701
20	0.928		0.714
21	0.898		0.727
22	0.855		0.740
23	0.819		0.752
24	0.796		0.763
25	0.770		0.774
26	0.754		0.785
27	0.706		0.796
28	0.696		0.806
29	0.673		0.815
30	0.639		0.825
31	0.622		0.834
32	0.588		0.842
33	0.570		0.851
34	0.545		0.858
35	0.538		0.866
36	0.519		0.874
37	0.494		0.881
38	0.471		0.888
39	0.459		0.894
40	0.443		0.901
41	0.430		0.907
42	0.409		0.913
43	0.401		0.919
44	0.378		0.924
45	0.352		0.929
46	0.345		0.934
47	0.336		0.939
48	0.311		0.944
49	0.300		0.948
50	0.284		0.952
51	0.276		0.956
52	0.258		0.960
53	0.251		0.964
54	0.240		0.967



"SKREE" ANALYSIS OF
PERCENT OF TOTAL VARIANCE

References

Bloom, B.S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, W.H. & Krathwohl, D.R. *Taxonomy of educational objectives: Handbook 1: Cognitive domain*. New York: David McKay, 1956.

Bruning, J.L., & Kintz, B.L. *Computational handbook of statistics.* Glenview, Illinois: Scott Foresman, 1968.

Cohen, J. *Statistical power analysis for the behavioral sciences.* New York: Academic Press, 1969.

Durnin, J. & Scandura, J.M. An algorithmic approach to assessing behavior potential: Comparison with item forms and hierarchical technologies. *Journal of Educational Psychology,* 1973, 65(2), 262-272.

Ebel, R.L. Estimation of the reliability of ratings. In W. Mehrens & R. Ebel (Eds.), *Principles of educational and psychological measurement,* Chicago: Rand McNally, 1967.

Frieze, I. & Weiner, B. Cue utilization and attributional judgements for success and failure. *Journal of Personality,* 1971, 39, 591-605.

Gozali, H., Cleary, T.A., Walster, G.W., & Gozali, J. Relationship between the internal-external control construct and achievement. *Journal of Educational Psychology,* 1973, 64, 15-21.

Haladyna, T.M. & Roid, G.H. The quality of domain-referenced test items. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April, 1976.

Harvey, J.H., Arkin, R.M., Gleason, J.M., & Johnston, S. Effect of expected and observed outcome of an action of the differential causal attributions of actor and observer. *Journal of Personality,* 1974, 42, 62-77.

Hays, W.L. *Statistics for psychologists*. New York: Holt, Rinehart, & Winston, 1963.

Henrysson, S. Gathering, analyzing, and using data on test items. In R.L. Thorndike (Ed.), *Educational measurement*. Washington, D.C.: American Council on Education, 1971.

Hively, W., Maxwell, G., Rabehl, G., Sension, D., & Lundin, S. *Domain-referenced curriculum evaluation: A technical handbook and a case study from the minniemast project*. Los Angeles, California: Center for the Study of Evaluation, University of California, Los Angeles, 1973.

Kelley, H.H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation*, 15. Lincoln: University of Nebraska Press, 1967.

Kohlberg, L. Early education: A cognitive-developmental view. *Child Development*, 1968, 39, 1013-1062.

Kukla, A. Attributional determinants of achievement-related behavior. *Journal of Personality and Social Psychology*, 1972, 21(2), 166-174.

Lefcourt, H.M. Internal versus external control of reinforcement: A review. *Psychological Bulletin*, 1966, 65(4), 206-220.

Lord, F.M. & Novick, M.R. *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley, 1968.

MacKenzie, K.D. *A behavioral constitution for structural change in little laboratory organizations* (Working paper no. 101). Lawrence, Kansas: School of Business, University of Kansas, 1975.

Program on Teaching Effectiveness. *Teacher training products: The state of the field* (Research and development memorandum no. 116). Stanford, California: Stanford Center for Research and Development in Teaching, Stanford University, 1974.

Robinson, J.P., Rush, J.G., & Head, K.B. Criteria for an attitude scale. In G.M. Maranell (Ed.), *Scaling: A sourcebook for behavioral scientists*. Chicago: Aldine Publishing Co., 1974.

Rotter, J.B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 1966, 80 (1, whole no. 609).

Rummel, R.J. *Applied factor analysis*. Evanston: Northwestern University Press, 1970.

Scriven, M.J. Evaluation perspectives and procedures. In W.J. Popham (Ed.), *Evaluation in education-current applications*. Berkeley: McCutchan Publishing Corporation, 1974.

Simon, J.G. & Feather, N.T. Causal attributions for success and failure at university examinations. *Journal of Educational Psychology*, 1973, 64, 57-62.

Stanley, J.C. Reliability. In R.L. Thorndike, *Educational measurement*. Washington, D.C.: American Council on Education, 1971.

Stayrook, N.G. & Corno, L. Student perceptions of teacher behavior as related to student achievement. Paper presented at the American Educational Research Association Annual Meeting, San Francisco, April, 1976.

Weiner, B. Attribution theory, achievement motivation, and the educational process. *Review of Educational Research*, 1972, 42, 203-215.

Weiner, B. & Kukla, A. An attributional analysis of achievement motivation. *Journal of Personality and Social Psychology*, 1970, 15(1), 1-20.